國家科學及技術委員會專題研究計畫申請書

申請條碼：113WIA0110259

一、基本資料：

計畫類別（單選）	一般研究計畫				
研究型別	個別型				
計畫歸屬	人文處				
申請機構/系所（單位）	中央研究院經濟研究所				
本計畫主持人姓名	蔡文禎	職稱	研究員	身分證號碼	G12055****
本計畫名稱 中文	結構性 Poisson 模型內含二元內生變數的一般化概似函數估計法				
本計畫名稱 英文	A generalized likelihood-based estimator for the structural Poisson model with K endogenous binary regressors				
整合型總計畫名稱					
整合型總計畫主持人				身分證號碼	
全程執行期限	自民國 113 年 08 月 01 日起至民國 115 年 07 月 31 日				
研究學門	學門代碼	學門名稱			
	H15A3	數理與數量方法			

【請考量己身負荷，申請適量計畫】
本年度申請主持本會各類研究計畫(含預核案)共 1 件。(共同主持之計畫不予計入)

本計畫是否同時有其他單位提供補助項目：■否； □是，請務必填寫表CM05*。

近三年內是否有執行非國科會補助之其他（含國內外、大陸地區及港澳）計畫：■否； □是，請務必填寫表CM14-1。

本計畫是否為國際合作研究：■否； □是，請加填表IM01~IM03

本計畫是否申請海洋研究船：■否； □是，請務必填寫表CM15

1.本計畫是否有進行下列實驗/研究：（勾選下列任一項，須附相關實驗/研究同意文件）
□人體試驗/人體檢體 □人類胚胎/人類胚胎幹細胞 □基因重組實驗 □基因轉殖田間試驗 □第二級以上感染性生物材料 □動物實驗(須同時加附動物實驗倫理3R說明)
2.本計畫是否為人文處行為科學研究計畫：□是(請檢附已送研究倫理審查之證明文件)； ■否
3.本計畫是否為人體試驗或人體研究計畫：□是(請增填研究中的性別考量檢核表CM16)； ■否

本計畫是否申請高效能計算資源：■否； □是，請務必填寫表CM17

計畫連絡人	姓名： 蔡文禎 電話：(公) 02-27822791-296 (宅/手機) 0936143806		
通訊地址	台北市南港研究院路二段128號中央研究院經濟研究所		
傳真號碼	02-27853946	E-MAIL	wtsay@econ.sinica.edu.tw

計畫主持人簽章：__________ 日期：__________

二、研究計畫中英文摘要：請就本計畫要點作一概述，並依本計畫性質自訂關鍵詞。

計畫中文關鍵詞	結構性 Poisson 模型，二元內生變數，概似函數估計法
計畫英文關鍵詞	Structural Poisson model, endogenous binary regressors, likelihood function
計畫中文摘要	本計畫研究結構性 Poisson 模型內含 k 個二元內生變數的一般化概似函數估計法的估計與檢定課題。因可同時處理 k 個內生二元自變數，本計畫所提出的方法是文獻上在此架構下最一般化的設定。預計將對經濟學產生許多可能應用，並可能擴及有其它學門的實證用途。 本計畫提出两種估計法，其一為两階段估計法，包含第一階段的多元 probit 估計以及第二階段的 Poisson 估計。其二將两階段合一成為一階段估計。两種方法的極限性質是計畫的重點之一，彼此間的效率比較也會在本計畫中探討。 為使計算有效率，除了導出概似函數的解析解外，概似函數的一階和二階微分值的解析解也會導出，以加速並準確估計模型的參數。本計畫也會以模擬實驗的方式驗証所提方法的小樣本表現。 本計畫也會進行實証研究，將探討多種两國間簽訂的合約，如最惠國待遇，雙邊租稅協定等四種內生二元變數對國際貿易額的影響。
計畫英文摘要	This project propose a most general structural Poisson model in the literature, because we allow k endogenous binary variables under this setup. This work is important, because the structural Poisson model has been intensively used in several fields of economics, including health, trade, investment and innovation. Two different QML estimators are proposed for implementing the estimation and inference procedure of the proposed model. We will also derive the analytical formulae of the score and Hessian functions of the log-likelihood functions of both the two-step QML and quasi-LIML estimators in order to estimate the parameters accurately and speed up the computation process greatly. We will derive the asymptotic distribution of the quasi-LIML estimator and that of the two-step QML one for statistical inference and conduct simulations to show the relative finite-sample performance of the proposed approaches. The relative theoretical efficiency between these estimators will be also investigated in this project. We conjecture that the quasi-LIML estimator is asymptotically more efficient than its two-step QML counterpart. We also apply our 2SPQML and quasi-LIML estimators to estimate the trade effect of four types of preferential economic integration agreements (PEIAs) empirically. This innovative approach will contribute to a comprehensive evaluation of the trade effects of PEIAs and provide useful insights for policymakers and researchers in the field.
計畫概述	請概述執行本計畫之目的及可能產生對人文、社會、經濟、學術發展等面向的預期影響性(三百字以內)。 ※此部分內容於獲核定補助後將逕予公開
	本計畫研究結構性 Poisson 模型內含 k 個二元內生變數的一般化概似函數估計法的估計與檢定課題。因可同時處理 k 個內生二元自變數，本計畫所提出的方法是文獻上在此架構下最一般化的設定。預計將對經濟學產生許多可能應用，並可能擴及有其它學門的實證用途。

III. Contents of Grant Proposal:

1. Research project's background:

This project proposes a novel approach to address the estimation and inference problems concerning the count data models with **endogenous binary** explanatory variables. These count models include an outcome equation with a structural-causal interpretation, along with additional equations that capture the generating processes of the endogenous binary variables (or treatment variables), hereafter referred as the generating process of binary endogenous explanatory variables (**BEEV**s). Essentially, the models mainly concern the impacts of a set of endogenous treatment variables on economic outcomes. Nevertheless, the explanatory variables in explaining the decision of each binary choice is also of interest to the literature, and will be investigated in this two-year project. In short, our methodology is to estimate a structural count model with BEEVs, instead of its reduced-form counterpart which does not pay attentions to the generating processes of these BEEVs. In the literature, structural count model has been considered intensively, including Terza (1998) and Wooldridge (2014).

The aforementioned structural count models with BEEVs can find extensive applications in various economic domains, such as health (Kenkel and Terza, 2001; Deb and Trivedi, 2006), innovation and investment (Hausman, Hall and Griliches, 1984; Head and Ries, 2008; Fahlenbrach, 2009), international trade (Santos Silva and Tenreyro, 2006; Egger et al., 2011), and the studies about the wage equations and production functions (Blackburn, 2007; Powell and Seabury, 2018). However, the endogeneity of binary explanatory variables in the structural count model is not completely tackled. For example, Terza (1998) develops a Poisson model that accommodates **only one** binary endogenous explanatory variable. Building on this research, Chen et al. (2023) expanded the capacity of such Poisson models to incorporate two BEEVs. In this project, we move a big step forward by providing an approach to estimate a Poisson models with flexible k BEEVs.[1]

Theoretically, we address the endogeneity of binary explanatory variables in the count models with a two-step Poisson quasi maximum likelihood (2SPQML) estimator. We favor the Poisson quasi-maximum likelihood (QML) over alternative estimators such as ordinary least squares (OLS) estimator or Poisson nonlinear least squares estimator, because the Poisson QML estimator remains consistent when the count data exhibiting the phenomenon of heteroskedasticity (Santos Silva and Tenreyro, 2006). Furthermore, the parameters of interest can be consistently estimated as long as the conditional mean of the response variable is correctly specified (White, 1982; Gourieroux, Monfort and Trognon, 1984), without relying on incidental distributional assumptions. Consequently, alternative estimators, such as the one designed for zero-inflated count data models by Mullahy (1986), are less appealing in this context due to their dependence on stronger assumptions about the parametric distribution of the data.[2] Interest-

[1] The instrumental variable (IV) estimation methods proposed by Mullahy (1997), Windmeijer and Silva (1997), and Jochmans and Verardi (2022) can handle BEEVs in count data models with fewer assumptions, while our approach is designed to deal with the structural Poisson model. Indeed, the choice of an econometric model should be contingent upon the specific economic question under consideration. For those interested in delving into the debate of IV versus structural models, please refer to the work of Heckman and Urzúa (2010).

[2] While many alternatives to Poisson regression within the domain of count data models

ingly, Santos Silva and Tenreyro (2006), Santos Silva and Tenreyro (2011), and Cohn, Liu and Wardlaw (2022) compare the performance of the Poisson QML estimator with various alternatives within the domain of count data models, finding that the Poisson QML estimator exhibits the most satisfactory performance in most cases.

This project also advances the understanding of the QML estimator by examining the relative asymptotic efficiency between the 2SPQML and the one-step joint estimator raised by Wooldridge (2014). Specifically, the outcome equation and the associated generating processes of BEEVs of the structural Poisson model can be either jointly estimated or accomplished through a two-step estimator such as the 2SPQML estimator considered in this project. For the one-step approach, this project adopts the quasi-Limited Information Maximum Likelihood (quasi-LIML) estimator of Wooldridge (2014). The feature of the quasi-LIML estimator is to maximize the sum of the two log-likelihood functions generated from conducting the first-step and the second-step QML for the structural Poisson model, and one or both of the log likelihoods may be misspecified. We aim to demonstrate, through both theoretical analysis and simulations, that the quasi-LIML

are more flexible, permitting overdispersion (see, e.g., Cameron and Trivedi (2013)), the concept of overdispersion lacks definition when the dependent variable lacks a natural scale. In cases where the dependent variable can be measured in diverse units, the relationship between the conditional mean and conditional variance becomes contingent on the scale of the data. Consequently, estimates derived from models allowing for overdispersion become sensitive to the scale of the dependent variable and the units in which it is measured, rendering them arbitrary. This issue, initially identified by Bosquet and Boulhol (2014) concerning the negative binomial estimator, extends to all estimators attempting to accommodate overdispersion, including zero-inflated models.

estimator is asymptotically more efficient than its two-step QML counterpart.[3]

Empirically, this project intends to apply our 2SPQML and quasi-LIML estimators to estimate the trade effect of preferential economic integration agreements (PEIAs). PEIAs, such as preferential trade agreements (PTAs), bilateral investment treaties (BITs), or double-taxation treaties (DTTs), serve as vital policy tools for countries to enhance trade and welfare. Consequently, assessing the effects of PEIAs on bilateral trade values has continued to be a central topic in international trade literature for over six decades, dating back to the work of Tinbergen (1962).

The caveat of our empirical studies hinges on the observation that the treatment of the endogeneity in PEIAs has been incomplete in the trade literature due to methodological constraints. For example, some studies either neglect the issue entirely or consider only one or two of these agreements. In particular, Egger et al. (2011) focused on endogenous PTA effects, while Chen et al. (2023) examined both endogenous PTA and BIT effects. In this project, we will enhance this understanding by distinguishing up to four types of PEIAs using the Design of Trade Agreements database. This nuanced approach will contribute to a comprehensive assessment of the trade effects of PEIAs and provide valuable insights for policymakers and researchers in the field.

This project consists of two year agenda. We present the details of each year's works, respectively.

[3]It is well-known that joint estimator is asymptotically more efficient than two-stage estimator if the objective function is the true log-likelihood, whereas a two-step estimator is more tractable and has lower computation cost.

The First Year

2. Methods, procedures, and implementation schedule

(1). Research principles, methods, and the innovation of research methods

Before addressing the way of taking care of the presence of endogenous binary explanatory variables in the count data models, we first present the econometric model, estimation strategy, and the accomplished works.

(1.1) Econometric model

When the response variable, y_1, is count data, it is popular to model y_1 using the exponential function:

$$\mathrm{E}[y_1|\mathbf{y}, \mathbf{x}, r_1] = \exp(\mathbf{x}_1\boldsymbol{\beta}_1 + r_1), \tag{1}$$

where $\mathbf{x}_1 = (1, \mathbf{x}, \mathbf{y})$ is a vector of explanatory variables, containing an intercept and any function of $\mathbf{y}$ and $\mathbf{x}$. $\boldsymbol{\beta}_1$ is a vector of parameters to be estimated and r_1 is the error term. In this context, we consider $\mathbf{x}$ to be a vector of exogenous variables and $\mathbf{y} = (y_2, y_3, \ldots, y_{k+1})$ denotes a $1 \times k$ vector of binary endogenous explanatory variables. Specifically, the elements in $\mathbf{x}$ are assumed to be independent of r_1, while $\mathbf{y}$ and r_1 are allowed to be correlated.

Extending the work of Terza (1998) and Chen et al. (2023), we assume the

BEEVs follow the following reduced form multivariate probit form:

$$
\begin{aligned}
y_j &= 1[\mathbf{z}\boldsymbol{\delta}_j + v_j \geq 0], \quad j = 2, 3, \ldots, k+1, && (2)\\
\mathrm{E}(v_j|\mathbf{z}) &= 0, && (3)\\
\mathrm{Var}(v_j|\mathbf{z}) &= 1, && (4)\\
\mathrm{Cov}(v_j, v_m|\mathbf{z}) &= \rho_{jm}, \quad m = j+1, \ldots, k+1, && (5)\\
\mathbf{v}|\mathbf{z} &\sim \mathrm{Normal}_k(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}}). && (6)
\end{aligned}
$$

Here, vector $\mathbf{z}$ includes a constant term, $\mathbf{x}$, and a vector of IVs. Thus $\mathrm{E}(r_1|\mathbf{z}) = 0$. $\mathbf{v} = (v_2, v_3, \ldots, v_{k+1})$ is a vector of error terms in Eq. (2). $\mathrm{Normal}_k(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}})$ denotes a k-variate standard normal distribution function with a mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{V}}$. $\boldsymbol{\delta_j}$ is a vector of parameters to be estimated.

To introduce endogeneity in $\mathbf{y}$, we assume the vector $\mathbf{r} = (r_1, \mathbf{v})$ follows a multivariate normal distribution with zero means and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{R}}$. Specifically, the covariance matrix is

$$
\boldsymbol{\Sigma}_{\mathbf{R}} = \begin{bmatrix} \alpha^2 & \boldsymbol{\rho} \\ \boldsymbol{\rho}^\top & \boldsymbol{\Sigma}_{\mathbf{V}} \end{bmatrix}. \tag{7}
$$

Note that the vector $\boldsymbol{\rho} = (\rho_2, \rho_3, \ldots, \rho_{k+1})$ contains covariance of r_1 and v_j, whereas parameter ρ_{jm} in Eq. (5) captures covariance between v_j and v_m.

(1.2) Two-step quasi-maximum likelihood estimation

Define $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^\top, \alpha, \boldsymbol{\rho})$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}_2^\top, \ldots, \boldsymbol{\delta}_{k+1}^\top, \boldsymbol{\Sigma}_{\mathbf{V}})$ as two sets of parameters to be estimated. Under the model assumptions outlined in Eqs. (2)–(6), the vector $\boldsymbol{\theta}_2$ can be consistently estimated using multivariate probit regression of $\mathbf{y}$ on $\mathbf{z}$. Specifically, for observed sample $(\mathbf{y}_i, \mathbf{z}_i)$, $i = 1, \ldots, n$, the first-step estimator, $\widehat{\boldsymbol{\theta}}_{2,2SPQML}$ maximizes the following objective function:

$$\sum_{i=1}^{n} q_{i2}(\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}_2) = \sum_{i=1}^{n} \log \Phi_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\star}_{\mathbf{V}}), \tag{8}$$

where

$$\begin{aligned} \boldsymbol{\mu} &\equiv (q_{i2}\mathbf{z}\boldsymbol{\delta}_2, \ldots, q_{ik+1}\mathbf{z}\boldsymbol{\delta}_{k+1}); & (9) \\ \boldsymbol{\Sigma}^{\star}_{\mathbf{V}} &\equiv \{\rho^{\star}_{jm}\} = \{q_{ij}q_{im}\rho_{jm}\}; & (10) \end{aligned}$$

$q_{ij} = 2y_{ij} - 1$; and Φ_k denotes the cumulative distribution function (CDF) of the standard multivariate normal distribution.[4]

The rest parameters of interest, $\boldsymbol{\theta}_1$, can be consistently estimated under suitable regularity conditions, resorting to Theorem 3.10 and Corollary 5.5 of White (1994). Different from the maximum likelihood method that specifies the conditional distribution of y_1, the QML method **approximates** it by a Poisson density function (White, 1994; Wooldridge, 2010):

$$f(y_1|\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{\exp(-\lambda^{\star}) \cdot (\lambda^{\star})^{y_1}}{y_1!}, \tag{11}$$

where $\lambda^{\star} = \mathrm{E}(y_1|\mathbf{y}, \mathbf{z})$ is a function of $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ to be derived.

Suppose we know the functional form of $\lambda^{\star}$. $\boldsymbol{\theta}_1$ can be consistently estimated by plug $\widehat{\boldsymbol{\theta}}_2$ in the second step of the Poisson quasi-maximum log-likelihood function:

$$\begin{aligned} \sum_{i=1}^{n} q_{i1}(y_1, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}_1, \widehat{\boldsymbol{\theta}}_2) &= \sum_{i=1}^{n} \ln f(y_{i1}|\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}_1, \widehat{\boldsymbol{\theta}}_2) \\ &= \sum_{i=1}^{n} \left[-\widehat{\lambda}^{\star}_i + y_{i1} \ln(\widehat{\lambda}^{\star}_i) - \ln(y_{i1}!) \right]. \end{aligned} \tag{12}$$

[4] Please refer to Chapter 17.5.7 of Greene (2012) for details on the multivariate probit model.

Here, $\widehat{\lambda}^\star = \lambda^\star(\boldsymbol{\theta}_1, \widehat{\boldsymbol{\theta}}_2)$. The 2SPQML estimator $\widehat{\boldsymbol{\theta}}_1$ solves

$$\max_{\boldsymbol{\vartheta}_1} \sum_{i=1}^{n} q_{i1}(y_1, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_1, \widehat{\boldsymbol{\theta}}_2). \tag{13}$$

Non-linear solvers in MATLAB's *fmincon* function are used to locate $\widehat{\boldsymbol{\theta}}_1$. However, this task is time-consuming if we rely on the solvers to calculate the score and Hessian of the quasi-log-likelihood function in Eq. (13) via numerical routines. We plan to resolve this numerical issue by providing the analytical formulae of the score and Hessian functions of the quasi-log-likelihood function.

(2) Anticipated problems and means of resolution

The key to implement the likelihood-based approach is to derive the value of $\lambda^\star$, and the gradient and Hessian matrices of the log-likelihood function. We first consider the way of computing $\lambda^\star$.

(2.1) Derivation of $\lambda^\star$

One of the bottlenecks of this research is to derive $\lambda^\star$. Here we present the key steps of the derivations. By definition, we have

$$\lambda^\star = \mathrm{E}(y_1 | \mathbf{y}, \mathbf{z}).$$

Applying the law of iterative expectation on the right-hand side, we obtain

$$\lambda^\star = \mathrm{E}\big[\mathrm{E}(y_1 | \mathbf{y}, \mathbf{x}, r_1) | \mathbf{y}, \mathbf{z}\big].$$

By applying Eq. (1), we replace $\mathrm{E}(y_1|\mathbf{y}, \mathbf{x}, r_1)$ on the right-hand side and arrive

$$\lambda^\star = \mathrm{E}\big[\exp(\mathbf{x}_1\boldsymbol{\beta}_1 + r_1) | \mathbf{y}, \mathbf{z}\big] = \exp(\mathbf{x}_1\boldsymbol{\beta}_1)\mathrm{E}\big[\exp(r_1) | \mathbf{y}, \mathbf{z}\big].$$

The conditional expectation term $\mathrm{E}\big[\exp(r_1)|\mathbf{y},\mathbf{z}\big]$ can be complicated. For example, in the bivariate normal case, i.e. $\mathbf{y}=(y_2,y_3)$, given the normal distribution assumption of the error terms, Eqs. (2)–(6), the above equation expands as:

$$\begin{aligned}\mathrm{E}[\exp(r_1)|y_2,y_3,\mathbf{z}] = {} & y_2y_3\mathrm{E}[\exp(r_1)|v_2>-\mathbf{z}\boldsymbol{\delta}_2, v_3>-\mathbf{z}\boldsymbol{\delta}_3] \\ & + y_2(1-y_3)\mathrm{E}[\exp(r_1)|v_2>-\mathbf{z}\boldsymbol{\delta}_2, v_3\le-\mathbf{z}\boldsymbol{\delta}_3] \\ & + (1-y_2)y_3\mathrm{E}[\exp(r_1)|v_2\le-\mathbf{z}\boldsymbol{\delta}_2, v_3>-\mathbf{z}\boldsymbol{\delta}_3] \\ & + (1-y_2)(1-y_3)\mathrm{E}[\exp(r_1)|v_2\le-\mathbf{z}\boldsymbol{\delta}_2, v_3\le-\mathbf{z}\boldsymbol{\delta}_3].\end{aligned} \tag{14}$$

In the k-variate normal case, there would be 2^k terms on the right-hand side of the conditional expectation, $\mathrm{E}[\exp(r_1)|\mathbf{v},\mathbf{z}]$. This task has never been tried and is not trivial. However, it is crucial for the implementation of the likelihood-based estimator for the structural Poisson model.

For clarity of illustration, we first show the way of evaluating the term in Eq. (14). Applying the law of iterative expectation on the right-hand side of Eq. (14), we have

$$\begin{aligned}\mathrm{E}[\exp(r_1)|y_2,y_3,\mathbf{z}] = {} & y_2y_3\mathrm{E}\big[\mathrm{E}[\exp(r_1)|v_2,v_3,\mathbf{z}]|v_2>-\mathbf{z}\boldsymbol{\delta}_2, v_3>-\mathbf{z}\boldsymbol{\delta}_3\big] \\ & + y_2(1-y_3)\mathrm{E}\big[\mathrm{E}[\exp(r_1)|v_2,v_3,\mathbf{z}]|v_2>-\mathbf{z}\boldsymbol{\delta}_2, v_3\le-\mathbf{z}\boldsymbol{\delta}_3\big] \\ & + (1-y_2)y_3\mathrm{E}\big[\mathrm{E}[\exp(r_1)|v_2,v_3,\mathbf{z}]|v_2\le-\mathbf{z}\boldsymbol{\delta}_2, v_3>-\mathbf{z}\boldsymbol{\delta}_3\big] \\ & + (1-y_2)(1-y_3)\mathrm{E}\big[\mathrm{E}[\exp(r_1)|v_2,v_3,\mathbf{z}]|v_2\le-\mathbf{z}\boldsymbol{\delta}_2, v_3\le-\mathbf{z}\boldsymbol{\delta}_3\big].\end{aligned} \tag{15}$$

It is clear now that the key step changes to derive the analytical functional form of $\mathrm{E}[\exp(r_1)|v_2,v_3,\mathbf{z}]$. This indicates that deriving $\mathrm{E}[\exp(r_1)|\mathbf{v},\mathbf{z}]$ is inevitable in k-variate normal case. Nevertheless, $\mathbf{z}$ is assumed to be independent of $(r_1,\mathbf{v})$, we realize that $\mathrm{E}[\exp(r_1)|\mathbf{v},\mathbf{z}] = \mathrm{E}[\exp(r_1)|\mathbf{v}]$. Accordingly, we need to derive the analytical functional form of $\mathrm{E}[\exp(r_1)|\mathbf{v}]$ for the subsequent analysis.

Before providing the way of evaluating $\mathrm{E}[\exp(r_1)|\mathbf{v}]$, we know that, by our multivariate normal assumption, the probability density functions (PDFs) for $\mathbf{r}$ and $\mathbf{v}$ are:

$$f_{\mathbf{R}}(\mathbf{r}) = \frac{\exp\left(-\mathbf{r}\mathbf{\Sigma}_{\mathbf{R}}^{-1}\mathbf{r}^{\top}/2\right)}{\sqrt{(2\pi)^{k+1}|\mathbf{\Sigma}_{\mathbf{R}}|}}, \tag{16}$$

and

$$f_{\mathbf{V}}(\mathbf{v}) = \frac{\exp\left(-\mathbf{v}\mathbf{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^{\top}/2\right)}{\sqrt{(2\pi)^{k}|\mathbf{\Sigma}_{\mathbf{V}}|}}. \tag{17}$$

Consider variable transformation $r_1^{\star} \equiv \exp(r_1)$ and the corresponding random vector $\mathbf{r}^{\star} \equiv (r_1^{\star}, \mathbf{v})$. The PDF of $\mathbf{r}^{\star}$ is:

$$f_{\mathbf{R}^{\star}}(\mathbf{r}^{\star}) = \frac{\exp\left[-\mathbf{r}'\mathbf{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^{\top}/2\right]}{\sqrt{(2\pi)^{k+1}|\mathbf{\Sigma}_{\mathbf{R}}|}} \cdot |\mathbf{J}|, \tag{18}$$

where $\mathbf{r}' = (\ln(r_1^{\star}), \mathbf{v})$, and $\mathbf{J}$ is the associated Jacobian matrix defined as

$$\mathbf{J} \equiv \begin{bmatrix} 1/r_1^{\star} & \mathbf{0}_{1,k} \\ \mathbf{0}_{k,1} & I_k \end{bmatrix}. \tag{19}$$

Here, $\mathbf{0}_{1,k} = \mathbf{0}_{k,1}^{\top}$ is a $1 \times k$ zero matrix and I_k is a $k \times k$ identity matrix. Clearly, $|\mathbf{J}| = 1/r_1^{\star}$.

By definition, the conditional PDF of $r_1^{\star}$ given $\mathbf{v}$ is given by:

$$\begin{aligned} g(r_1^{\star}|\mathbf{v}) &= \frac{f_{\mathbf{R}^{\star}}(\mathbf{r}^{\star})}{f_{\mathbf{V}}(\mathbf{v})} \\ &= \frac{1}{r_1^{\star}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{|\mathbf{\Sigma}_{\mathbf{R}}|/|\mathbf{\Sigma}_{\mathbf{V}}|}} \cdot \exp\left[-\frac{\mathbf{r}'\mathbf{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^{\top} - \mathbf{v}\mathbf{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^{\top}}{2}\right]. \end{aligned} \tag{20}$$

We aim to demonstrate that the conditional PDF, $g(r_1^{\star}|\mathbf{v})$, follows a log-normal distribution. Thus, we can derive the analytical expression for $\mathrm{E}[\exp(r_1)|\mathbf{v}]$.

We first assume that $\boldsymbol{\Sigma}_{\mathbf{V}}$ is invertible, and then we introduce the Schur complement of $\Sigma_{\mathbf{V}}$ in covariance matrix $\boldsymbol{\Sigma}_{\mathbf{R}}$(which is defined in Eq. (7)):

$$\sigma_{r_1^\star}^2 \equiv \boldsymbol{\Sigma}_{\mathbf{R}}/\boldsymbol{\Sigma}_{\mathbf{V}} = \alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top. \tag{21}$$

According to Theorem 1.1 (Schur's Formula) in Horn and Zhang (2005), we have

$$|\boldsymbol{\Sigma}_{\mathbf{R}}|/|\boldsymbol{\Sigma}_{\mathbf{V}}| = |\boldsymbol{\Sigma}_{\mathbf{R}}/\boldsymbol{\Sigma}_{\mathbf{V}}| = |\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top|.$$

Moreover, since covariance matrix $\boldsymbol{\Sigma}_{\mathbf{R}}$ is positive semi-definite, Theorem 1.20 in Horn and Zhang (2005) ensures that $\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top \geq 0$. This establishes that

$$|\boldsymbol{\Sigma}_{\mathbf{R}}|/|\boldsymbol{\Sigma}_{\mathbf{V}}| = |\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top| = \alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top. \tag{22}$$

Thus we can simplify the conditional PDF of $r_1^\star$ given $\mathbf{v}$ to

$$g(r_1^\star|\mathbf{v}) = \frac{1}{r_1^\star} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top}} \cdot \exp\left[-\frac{\mathbf{r}'\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^\top - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^\top}{2}\right]. \tag{23}$$

And the next step is to evaluate the term, $\mathbf{r}'\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^\top - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^\top$, in this conditional PDF.

According to Theorem 1.2 in Horn and Zhang (2005), we can obtain $\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}$:

$$\boldsymbol{\Sigma}_{\mathbf{R}}^{-1} = \begin{bmatrix} \sigma_{r_1^\star}^{-2} & -\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} \\ -\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2} & \boldsymbol{\Sigma}_{\mathbf{V}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} \end{bmatrix}, \tag{24}$$

where $\sigma_{r_1^\star}$ is defined in Eq. (21). Using this result, we have:

$$\begin{aligned}
\mathbf{r}'\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^\top &= \mathbf{r}'\begin{bmatrix} \sigma_{r_1^\star}^{-2} & -\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} \\ -\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2} & \boldsymbol{\Sigma}_{\mathbf{V}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} \end{bmatrix}(\mathbf{r}')^\top \\
&= \left[\ln(r_1^\star)\sigma_{r_1^\star}^{-2} - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2} \quad -\ln(r_1^\star)\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} + \mathbf{v}(\boldsymbol{\Sigma}_{\mathbf{V}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1})\right](\mathbf{r}')^\top \\
&= \ln(r_1^\star)^2\sigma_{r_1^\star}^{-2} - 2\ln(r_1^\star)\mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2} + \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^\top + \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top\sigma_{r_1^\star}^{-2}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^\top.
\end{aligned} \tag{25}$$

Thus the term $\mathbf{r}'\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^{\top} - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^{\top}$ can be simplified as:

$$
\begin{aligned}
\mathbf{r}'\boldsymbol{\Sigma}_{\mathbf{R}}^{-1}(\mathbf{r}')^{\top} - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^{\top} &= \sigma_{r_1^\star}^{-2}\left[\ln(r_1^\star)^2 - 2\ln(r_1^\star)\mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^{\top} + \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^{\top}\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^{\top}\right] \\
&= \sigma_{r_1^\star}^{-2}\left[\ln(r_1^\star) - \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^{\top}\right]^2.
\end{aligned}
\tag{26}
$$

Finally, we introduce the definition:

$$
\mu_{r_1^\star} \equiv \mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^{\top}. \tag{27}
$$

Subsequently, we can simplify the conditional PDF of $r_1^\star$ given $\mathbf{v}$ to

$$
g(r_1^\star|\mathbf{v}) = \frac{1}{r_1^\star}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma_{r_1^\star}}\cdot\exp\left(-\frac{\left[\ln(r_1^\star) - \mu_{r_1^\star}\right]^2}{2\sigma_{r_1^\star}^2}\right). \tag{28}
$$

It follows that the conditional distribution of $r_1^\star$ is log-normal distribution with conditional mean and variance:

$$
\begin{aligned}
\mathrm{E}[\exp(r_1)|\mathbf{v}] &= \exp(\mu_{r_1^\star} + \sigma_{r_1^\star}^2/2), &(29)\\
\mathrm{Var}[\exp(r_1)|\mathbf{v}] &= [\sigma_{r_1^\star}^2 - 1]\exp(2\mu_{r_1^\star} + \sigma_{r_1^\star}^2). &(30)
\end{aligned}
$$

Having this result allows us to derive $\mathrm{E}[\exp(r_1)|\mathbf{y},\mathbf{z}]$. For the ease of exposition, we apply Eq. (29) to replace $\mathrm{E}[\exp(r_1)|\mathbf{z}, v_2, v_3]$ in our bivariate example Eq. (15):

$$
\begin{aligned}
\mathrm{E}[\exp(r_1)|y_2, y_3, \mathbf{z}] = {} & y_2y_3\mathrm{E}\big[\exp(\mu_{r_1^\star} + \sigma_{r_1^\star}^2/2)|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3\big] \\
& + y_2(1-y_3)\mathrm{E}\big[\exp(\mu_{r_1^\star} + \sigma_{r_1^\star}^2/2)|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 \le -\mathbf{z}\boldsymbol{\delta}_3\big] \\
& + (1-y_2)y_3\mathrm{E}\big[\exp(\mu_{r_1^\star} + \sigma_{r_1^\star}^2/2)|v_2 \le -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3\big] \\
& + (1-y_2)(1-y_3)\mathrm{E}\big[\exp(\mu_{r_1^\star} + \sigma_{r_1^\star}^2/2)|v_2 \le -\mathbf{z}\boldsymbol{\delta}_2, v_3 \le -\mathbf{z}\boldsymbol{\delta}_3\big].
\end{aligned}
$$

Since $\sigma_{r_1^\star}$ is a function of parameters, we factor out $\exp(\sigma_{r_1^\star}^2/2) = \exp[(\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top)/2]$ in the above equation and get:

$$
\begin{aligned}
\mathrm{E}[\exp(r_1)|y_2, y_3, \mathbf{z}] = \exp[(\alpha^2 - \boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top)/2]\times \\
&\big\{y_2 y_3 \mathrm{E}[\exp(\mu_{r_1^\star})|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ y_2(1-y_3)\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 \leq -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ (1-y_2)y_3\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 \leq -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ (1-y_2)(1-y_3)\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 \leq -\mathbf{z}\boldsymbol{\delta}_2, v_3 \leq -\mathbf{z}\boldsymbol{\delta}_3]\big\}.
\end{aligned} \tag{31}
$$

Considering that we cannot identify parameter α, let the value of $\boldsymbol{\beta}_1^*$ be $\boldsymbol{\beta}_1$ with the first element plus an extra $\alpha^2/2$. Therefore, we have

$$
\mathrm{E}[y_1|y_2, y_3, \mathbf{z}] = \exp(\mathbf{x}_1\boldsymbol{\beta}_1)\mathrm{E}[\exp(r_1)|y_2, y_3, \mathbf{z}] = \exp(\mathbf{x}_1\boldsymbol{\beta}_1^*)\Psi(\boldsymbol{\rho}, \boldsymbol{\Sigma}_{\mathbf{V}}, \boldsymbol{\delta}), \tag{32}
$$

where

$$
\begin{aligned}
\Psi(\mathbf{v}, \mathbf{z}; \boldsymbol{\rho}, \boldsymbol{\Sigma}_{\mathbf{V}}, \boldsymbol{\delta}) = \exp(-\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top/2)\times \\
&\big\{y_2 y_3 \mathrm{E}[\exp(\mu_{r_1^\star})|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ y_2(1-y_3)\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 > -\mathbf{z}\boldsymbol{\delta}_2, v_3 \leq -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ (1-y_2)y_3\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 \leq -\mathbf{z}\boldsymbol{\delta}_2, v_3 > -\mathbf{z}\boldsymbol{\delta}_3] \\
&+ (1-y_2)(1-y_3)\mathrm{E}[\exp(\mu_{r_1^\star})|v_2 \leq -\mathbf{z}\boldsymbol{\delta}_2, v_3 \leq -\mathbf{z}\boldsymbol{\delta}_3]\big\}.
\end{aligned} \tag{33}
$$

Now, the next step of deriving the analytical form for $\lambda^\star = \mathrm{E}(y_1|\mathbf{y}, \mathbf{z})$ is to obtain the functional form of Ψ for the bivariate case.

Let's denote the standard bivariate normal PDF and the standard bivariate normal CDF as $\phi_2(a, b, \rho)$ and $\Phi_2(a, b, \rho)$, respectively, where $a, b \in \mathbb{R}$, and $\rho \in$

$[-1, 1]$. The four conditional expectations in Eq. (33) are:

$$
\begin{aligned}
&\mathbf{E}[\exp(\mu_{r_1^\star})|v_2 > -\boldsymbol{z}\boldsymbol{\delta}_2, v_3 > -\boldsymbol{z}\boldsymbol{\delta}_3, \boldsymbol{z}] \\
&\quad = \int_{-\boldsymbol{z}\boldsymbol{\delta}_2}^{\infty}\int_{-\boldsymbol{z}\boldsymbol{\delta}_3}^{\infty} \exp(\mu_{r_1^\star})\frac{\phi_2(v_2, v_3, \rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})} dv_2 dv_3 \\
&\quad = \exp\left[\frac{\rho_2^2 + \rho_3^2 - 2\rho_{23}\rho_2\rho_3}{2(1-\rho_{23}^2)}\right] \frac{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2 + \rho_2, \boldsymbol{z}\boldsymbol{\delta}_3 + \rho_3, \rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})},
\end{aligned} \tag{34}
$$

$$
\begin{aligned}
&\mathbf{E}[\exp(\mu_{r_1^\star})|v_2 > -\boldsymbol{z}\boldsymbol{\delta}_2, v_3 \le -\boldsymbol{z}\boldsymbol{\delta}_3, \boldsymbol{z}] \\
&\quad = \int_{-\boldsymbol{z}\boldsymbol{\delta}_2}^{\infty}\int_{-\infty}^{-\boldsymbol{z}\boldsymbol{\delta}_3} \exp(\mu_{r_1^\star})\frac{\phi_2(v_2, v_3, -\rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})} dv_2 dv_3 \\
&\quad = \exp\left[\frac{\rho_2^2 + \rho_3^2 - 2\rho_{23}\rho_2\rho_3}{2(1-\rho_{23}^2)}\right] \frac{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2 + \rho_2, -\boldsymbol{z}\boldsymbol{\delta}_3 - \rho_3, -\rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})},
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
&\mathbf{E}[\exp(\mu_{r_1^\star})|v_2 \le -\boldsymbol{z}\boldsymbol{\delta}_2, v_3 > -\boldsymbol{z}\boldsymbol{\delta}_3, \boldsymbol{z}] \\
&\quad = \int_{-\infty}^{-\boldsymbol{z}\boldsymbol{\delta}_2}\int_{-\boldsymbol{z}\boldsymbol{\delta}_3}^{\infty} \exp(\mu_{r_1^\star})\frac{\phi_2(v_2, v_3, -\rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})} dv_2 dv_3 \\
&\quad = \exp\left[\frac{\rho_2^2 + \rho_3^2 - 2\rho_{23}\rho_2\rho_3}{2(1-\rho_{23}^2)}\right] \frac{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2 - \rho_2, \boldsymbol{z}\boldsymbol{\delta}_3 + \rho_3, -\rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})},
\end{aligned} \tag{36}
$$

and

$$
\begin{aligned}
&\mathbf{E}[\exp(\mu_{r_1^\star})|v_2 \le -\boldsymbol{z}\boldsymbol{\delta}_2, v_3 \le -\boldsymbol{z}\boldsymbol{\delta}_3, \boldsymbol{z}] \\
&\quad = \int_{-\infty}^{-\boldsymbol{z}\boldsymbol{\delta}_2}\int_{-\infty}^{-\boldsymbol{z}\boldsymbol{\delta}_3} \exp(\mu_{r_1^\star})\frac{\phi_2(v_2, v_3, \rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})} dv_2 dv_3 \\
&\quad = \exp\left[\frac{\rho_2^2 + \rho_3^2 - 2\rho_{23}\rho_2\rho_3}{2(1-\rho_{23}^2)}\right] \frac{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2 - \rho_2, -\boldsymbol{z}\boldsymbol{\delta}_3 - \rho_3, \rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})}.
\end{aligned} \tag{37}
$$

The second equalities in each of the aforementioned four conditional expectations, Eqs. (34)–(37), are based on the results of Eq. (E.1) in Chen et al. (2023):

$$
\begin{aligned}
&\exp(\mu_{r_1^\star})\phi_2(v_2, v_3, \rho_{23}) \\
&\quad = \exp\left[\frac{\rho_2^2 + \rho_3^2 - 2\rho_{23}\rho_2\rho_3}{2(1-\rho_{23}^2)}\right] \phi_2(v_2 + \rho_2, v_3 + \rho_3, \rho_{23}).
\end{aligned}
$$

We now can write Ψ in Eq. (33) as:

$$\begin{aligned}\Psi(\rho_2, \rho_3, \rho_{23}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) = {} & y_2 y_3 \frac{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2 + \rho_2, \boldsymbol{z}\boldsymbol{\delta}_3 + \rho_3, \rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})} \\ & + y_2(1-y_3)\frac{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2 + \rho_2, -\boldsymbol{z}\boldsymbol{\delta}_3 - \rho_3, -\rho_{23})}{\Phi_2(\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})} \\ & + (1-y_2)y_3\frac{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2 - \rho_2, \boldsymbol{z}\boldsymbol{\delta}_3 + \rho_3, -\rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, \boldsymbol{z}\boldsymbol{\delta}_3, -\rho_{23})} \\ & + (1-y_2)(1-y_3)\frac{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2 - \rho_2, -\boldsymbol{z}\boldsymbol{\delta}_3 - \rho_3, \rho_{23})}{\Phi_2(-\boldsymbol{z}\boldsymbol{\delta}_2, -\boldsymbol{z}\boldsymbol{\delta}_3, \rho_{23})}.\end{aligned}$$

Moreover, denoting $q_2 = 2y_2 - 1$, $q_3 = 2y_3 - 1$, $\mu_2 = q_2\boldsymbol{z}\boldsymbol{\delta}_2$, $\mu_3 = q_3\boldsymbol{z}\boldsymbol{\delta}_3$, and $\boldsymbol{\rho}^\star = q_2 q_3 \rho_{23}$, we have a much simpler form:

$$\Psi(\rho_2, \rho_3, \rho_{23}, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) = \frac{\Phi_2(\mu_2 + q_2\rho_2, \mu_3 + q_3\rho_3, \boldsymbol{\rho}^\star)}{\Phi_2(\mu_2, \mu_3, \boldsymbol{\rho}^\star)}. \tag{38}$$

We can use similar steps to derive Ψ of the k-variate case with one important modification. In the aforementioned bivariate case, there are $2^2 = 4$ possible combinations of y_2 and y_3 values, resulting in four terms in Eq. (33). It follows that in the k-variate case, there will be 2^k possible combination of $\mathbf{y}$ values. To summarize these 2^k possible combinations, we follow the work of Chib and Greenberg (1998) and write a general form of Eq. (33) as

$$\Psi(\mathbf{v}, \mathbf{z}; \boldsymbol{\rho}, \boldsymbol{\Sigma}_{\mathbf{V}}, \boldsymbol{\delta}) = \exp\left(-\frac{\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top}{2}\right)\int_{A_{k+1}}\cdots\int_{A_2}\exp(\mu_{r_1^\star})\cdot\frac{\phi_k(\mathbf{v}|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}})}{\Phi_k(\boldsymbol{\mu}|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}}^\star)}d\mathbf{v}, \tag{39}$$

where ϕ_k and Φ_k are the density and distribution of a k-variate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{V}}$. Note that the vector $\boldsymbol{\mu}$ and matrix $\boldsymbol{\Sigma}_{\mathbf{V}}^\star$ are defined in Eqs. (9)–(10). Set A_j to be the interval:

$$A_j = \begin{cases}(-\infty, \mathbf{z}\boldsymbol{\delta}_j) & \text{if } y_j = 1, \\ (\mathbf{z}\boldsymbol{\delta}_j, \infty) & \text{if } y_j = 0,\end{cases} \tag{40}$$

for $j = 2, \ldots, k+1$.

We show that the products of these three terms, $\exp(-\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top/2)$, $\exp(\mu_{r_1^\star})$, and $\phi_k(\mathbf{v}|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}})$ in Eq. (39) is the PDF of a k-variate normal distribution with mean $\boldsymbol{\rho}$ and variance $\boldsymbol{\Sigma}_{\mathbf{V}}$:

$$\begin{aligned}
&\exp\left(-\frac{\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top}{2}\right)\exp(\mu_{r_1^\star})\phi_k(\mathbf{v}|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}}) \\
&\qquad = \exp\left(-\frac{\boldsymbol{\rho}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top}{2}\right)\exp(\mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\rho}^\top)\cdot\frac{\exp(-\mathbf{v}\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{v}^\top/2)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}_{\mathbf{V}}|}} \\
&\qquad = \frac{\exp[-(\mathbf{v}-\boldsymbol{\rho})\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}(\mathbf{v}-\boldsymbol{\rho})^\top/2]}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}_{\mathbf{V}}|}}.
\end{aligned} \tag{41}$$

Thus we can further simplify Eq. (39) as

$$\Psi(\mathbf{v}, \mathbf{z}; \boldsymbol{\rho}, \boldsymbol{\Sigma}_{\mathbf{V}}, \boldsymbol{\delta}) = \frac{\Phi_k(\boldsymbol{\mu}+\boldsymbol{\rho}^\star|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}}^\star)}{\Phi_k(\boldsymbol{\mu}|\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{V}}^\star)}, \tag{42}$$

where $\boldsymbol{\rho}^\star = \mathbf{q}\odot\boldsymbol{\rho}$ and $\mathbf{q} = (q_2, \ldots, q_{k+1})$. The notation $\odot$ is Hadamard product or element-wise product. This formula is equivalent to the bivariate case of Eq. (38) in which:

$$\begin{aligned}
\boldsymbol{\mu} &= (q_2\mathbf{z}\boldsymbol{\delta}_2, q_3\mathbf{z}\boldsymbol{\delta}_3), \\
\boldsymbol{\rho}^\star &= (q_2\rho_2, q_3\rho_3), \\
\boldsymbol{\Sigma}_{\mathbf{V}}^\star &= \begin{bmatrix} 1 & q_2q_3\rho_{23} \\ q_2q_3\rho_{23} & 1 \end{bmatrix}.
\end{aligned}$$

As a result, we derive the analytic form of $\lambda^\star$ as

$$\lambda^\star = \exp(\mathbf{x}_1\boldsymbol{\beta}_1^\star)\Psi. \tag{43}$$

(2.2) Works planned to be done in the first year project

The focus of the first year project is to develop a 2SPQML estimator that addresses the endogeneity of binary explanatory variables in the structural count data model. We show in Section 2.1 how to obtain this 2SPQML estimator by deriving the analytical expression of $\lambda^{\star}$. Based on the findings in Section 2.1, we are close to achieve this goal once we obtain the analytical expression of Eq. (39).

However, using MATLAB's non-linear solvers to numerically compute the score and Hessian of the quasi-log-likelihood function would be very inefficient. We plan to overcome this numerical challenge by further deriving the analytical formulae of the score and Hessian functions of the log-likelihood function. This work needs a lot of analysis studies, but is crucial for the implementation of the approach proposed in this project. Moreover, we will derive the asymptotic distribution of the 2SPQML estimator for statistical inference.

Finally, we will design appropriate simulations to evaluate the performance of the 2SPQML estimator. Then, we will compare its performance to the conventional Poisson QML estimator.

Specifically, the Monte Carlo simulation is designed for the different sample sizes, $n \in \{200^2, 400^2, 800^2\}$, under the following data generating process (DGP):

$$y_{1,ij} = \exp(\beta_{11} + \beta_{12}x_{12,ij} + \beta_{13}y_{2,ij} + \beta_{14}y_{3,ij} + r_{1,ij}), \tag{44}$$

$$y_{2,ij} = 1[\delta_{21} + \delta_{22}z_{1,ij} + \delta_{23}z_{2,ij} + \delta_{24}z_{3,ij} + v_{2,ij} \geq 0], \tag{45}$$

$$y_{3,ij} = 1[\delta_{31} + \delta_{32}z_{1,ij} + \delta_{33}z_{2,ij} + \delta_{34}z_{3,ij} + v_{3,ij} \geq 0], \tag{46}$$

$$y_{4,ij} = 1[\delta_{41} + \delta_{42}z_{1,ij} + \delta_{43}z_{2,ij} + \delta_{44}z_{3,ij} + v_{4,ij} \geq 0], \tag{47}$$

where i, j = 1,2,..., and up to 200, 400, or 800. $(r_{1,ij}, v_{2,ij}, v_{3,ij}, v_{4,ij})$ are drawn from a multivariate normal distribution with mean zero and covariance, $\mathbf{\Sigma}$, as:

$$\mathbf{\Sigma} = \begin{bmatrix} \alpha^2 & \mu_2 & \mu_3 & \mu_4 \\ \mu_2 & 1 & \rho_{23} & \rho_{24} \\ \mu_3 & \rho_{23} & 1 & \rho_{34} \\ \mu_4 & \rho_{24} & \rho_{34} & 1 \end{bmatrix}. \tag{48}$$

We note that, in the spirit of "quasi"-maximum likelihood, the dependent variable $y_{1,ij}$ is not generated as count data (i.e., non-negative integer) nor follows a Poisson process.

The covariates $x_{12,ij}, z_{1,ij}, z_{2,ij}$, and $z_{3,ij}$ are all drawn independently from uniform distribution $U(-0.5, 0.5)$, and the parameters of interest are set by

- $(\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}) = (-1, 0, 1, 1)$
- $(\delta_{21}, \delta_{22}, \delta_{23}, \delta_{24}) = (0, 0.5, -1, 1)$
- $(\delta_{31}, \delta_{32}, \delta_{33}, \delta_{34}) = (0, 0.5, 1, -1)$
- $(\delta_{41}, \delta_{42}, \delta_{43}, \delta_{44}) = (1, -1, 1, -1)$

Moreover, we use the estimates from the conventional Poisson QML estimator of β_{12}, β_{13}, and β_{14} and add them with the extra random numbers generated from Uniform$(-0.5, 0.5)$ to serve as their initial values for implementing the proposed method and to create a realistic simulation scheme. As for the initial values for estimating μ_2, μ_3, and μ_4, we simply use a random vector generated from Uniform$(-0.5, 0.5)$ as their starting values.

We evaluate the performance of the Poisson 2SPQML estimator under three different endogeneity levels: $\mu_2 = \mu_3 = \mu_4 = \rho_{23} = \rho_{24} = \rho_{34} = \xi$, where

$\xi \in \{0.3, 0.6, 0.9\}$. We compare the average biases, and root-mean-squared errors (RMSE) of the Poisson 2SPQML estimator with those of the conventional Poisson QML method, which treats all covariates as exogenous variables. For ease of exposition, we refer to the conventional Poisson QML as Poisson QML for the rest of the project. The number of replications for each configuration of chosen parameters is 1,000.

3. Anticipated results and achievements:

- Derive the analytical formulae of the score and Hessian functions of the log-likelihood function.
- Derive the asymptotic distribution of the 2SPQML estimator for statistical inference and conduct simulations to show the finite-sample performance of the proposed approach.
- Submit the report to international journal for publication to spread out the knowledge thus obtained.
- The assistant and the principle investigator of this project will get a fruitful knowledge about the proposed likelihood-based method for the generalized structural Poisson model.
- The assistant will obtain solid training in coding, writing a paper, and collecting data.
- We expect the project to be published in journal like *Journal of Econometrics*, *Journal of Applied Econometrics*, or *Review of Economics and Statistics*, or other related journals.

The Second Year

The objective of the second year project is to advance the understanding of the QML estimator by proposing one-step joint estimator as compared to the two-step approach considered in the first year project. we will examine the relative asymptotic efficiency of the one-step and two-step QML estimators for the generalized structural Poisson model by analysis and Monte Carlo simulations. We will also apply the method to an empirical study about international trade.

2. Methods, procedures, and implementation schedule:

(1). Research principles, methods, and the innovation of research methods

(1.1) Quasi-limited information maximum likelihood estimation

Instead of estimating $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ separately, we can estimate them in one step using the quasi-LIML estimator proposed by Wooldridge (2014). The quasi-LIML estimator $(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2)$ is solution to the following objective function:

$$\max_{\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2} \sum_{i=1}^{n} [q_{i1}(y_1, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) + q_{i2}(\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_2)]. \tag{49}$$

The one-step quasi-LIML estimator has never been considered for the generalized structural Poisson model. We conjecture this one-step approach will be more efficient than its two-step counterpart considered in our first year project. Therefore, we will derive the asymptotic distribution of the proposed one-step method to complete this efficiency comparison. Again, this work has never been considered in the literature. Moreover, the estimation and inference procedures for the one-step quasi-LIML estimator will also be the major works of this second year project.

(2) Anticipated problems and means of resolution

(2.1) Works planned to be done in the second year project

As discussed previously, the asymptotic distribution of the quasi-LIML estimator will be derived. This work is technically intensive, and will be an important contribution of this second year project.

For the estimation issue, we can use MATLAB's solvers to find the values for $(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2)$ from the objective function in Eq. (49). This task is even more time-consuming for the LIML case than its two-step counterpart. Therefore, we cannot rely on the solvers to calculate the score and Hessian of the quasi-log-likelihood function via numerical routines, and we need to derive the value of the gradient and Hessian matrices of the log-likelihood function in Eq. (49) in order to improve the computational speed. As compared to the works planned for our first year project, the analysis works are even more difficult for the second year project.

Finally, we will design appropriate simulations to evaluate the relative finite-sample performance between the 2SPQML estimator and quasi-LIMI estimators. The second Monte Carlo simulation of this two-year project is designed to demonstrate the performance of the QML estimators for the widely known gravity model in international economics. We use $n = 2,500$, $10,000$, and $22,500$ to align with the regular sample size observed in international economics. For example, Egger et al. (2011) include $126 \times 125 = 15,750$ country-pairs in their dataset.

We also include two sets of dummies to imitate the scenario of a gravity model with two-way fixed effects along with the proceeding endogeneity setting in the error terms, $(r_{1,ij}, v_{2,ij}, v_{3,ij},, v_{4,ij})$. The new DGP thus becomes:

$$y_{1,ij} = \exp(\beta_{11} + \beta_{12}x_{12,ij} + \beta_{13}y_{2,ij} + \beta_{14}y_{3,ij} + e_{1,i} + m_{1,j} + r_{1,ij}), \tag{50}$$

$$y_{2,ij} = 1[\delta_{21} + \delta_{22}z_{1,ij} + \delta_{23}z_{2,ij} + \delta_{24}z_{3,ij} + e_{2,i} + m_{2,j} + v_{2,ij} \geq 0], \tag{51}$$

$$y_{3,ij} = 1[\delta_{31} + \delta_{32}z_{1,ij} + \delta_{33}z_{2,ij} + \delta_{34}z_{3,ij} + e_{3,i} + m_{3,j} + v_{3,ij} \geq 0], \tag{52}$$

$$y_{4,ij} = 1[\delta_{41} + \delta_{42}z_{1,ij} + \delta_{43}z_{2,ij} + \delta_{44}z_{3,ij} + e_{4,i} + m_{4,j} + v_{4,ij} \geq 0], \tag{53}$$

where the terms, $e_{j,i}$ and $m_{i,j}$, represent the two-way fixed effects included in the above four equations. Therefore, the computation burdens of this simulation design is much higher than the one in the first year project. The letter 'e' denotes the meaning of exporter, while the letter 'm' conveys the idea of an importer. Moreover, these eight terms are independently generated from the random variables O_1, where $O_1 \sim \text{Uniform}(-0.5, 0.5)$ across all i and j considered in this project.

Since the simulation design includes intra-national trade country pairs, the number of included importer and exporter dummies in Eqs. (50)–(53) is $2 \times N$, where N denotes the country number.

In this project, we apply our proposed method to the dataset used in Egger et al. (2011) where they only consider the case with a single endogenous binary regressor. Egger et al. (2011) are interested in evaluating the effects of signing PTAs on the magnitude of bilateral trade flows.

We plan to apply our 2SPQML and quasi-LIML estimators to estimate the trade effect of preferential economic integration agreements (PEIAs). This not only extends the coverage of the literature, but also helps us check the robustness of the findings in Egger et al. (2011). The dataset of Egger et al. (2011) includes the information concerning 126 countries for the year 2005. We thus have $126 \times 125 = 15{,}750$ bilateral trade country-pairs. In this research, we aim to improve this literature by distinguishing up to four types of PEIAs using the Design of Trade Agreements database. Furthermore, the importer and exporter fixed effects are both included with the other country specific determinants such as a country's GDP, population, and capital-labor ratio in the model specifications. This refined approach will contribute to a comprehensive evaluation of the trade effects of PEIAs and provide useful insights for policymakers and researchers in the field.

3. Anticipated results and achievements

- Derive the analytical formulae of the score and Hessian functions of the log-likelihood function of the quasi-LIML estimator.
- Derive the asymptotic distribution of the quasi-LIML estimator for statistical inference and conduct simulations to show the relative finite-sample performance of the proposed approaches.
- Conduct empirical study to show the usefulness of the proposed approach.
- Submit the report to international journal for publication to spread out the knowledge thus obtained.

- The assistant and the principle investigator of this project will get a fruitful knowledge about the proposed likelihood-based method for the generalized structural Poisson model.

- We expect the project to be published in journal like *Journal of Econometrics*, *Journal of Applied Econometrics*, or *Review of Economics and Statistics*, or other related journals.

4. Integrated research project

This project propose a most general structural Poisson model in the literature, because we allow k endogenous binary variables under this setup. This work is important, because the structural Poisson model has been intensively used in several fields of economics, including health, trade, investment and innovation.

Two different QML estimators are proposed for implementing the estimation and inference procedure of the proposed model. We will also derive the analytical formulae of the score and Hessian functions of the log-likelihood functions of both the two-step QML and quasi-LIML estimators in order to estimate the parameters accurately and speed up the computation process greatly.

We will derive the asymptotic distribution of the quasi-LIML estimator and that of the two-step QML one for statistical inference and conduct simulations to show the relative finite-sample performance of the proposed approaches. The relative theoretical efficiency between these estimators will be also investigated in this project. We conjecture that the quasi-LIML estimator is asymptotically more efficient than its two-step QML counterpart.

We also apply our 2SPQML and quasi-LIML estimators to estimate the trade effect of four types of preferential economic integration agreements (PEIAs) empirically. This innovative approach will contribute to a comprehensive evaluation of the trade effects of PEIAs and provide useful insights for policymakers and researchers in the field.

References

Blackburn, McKinley L. 2007. "Estimating Wage Differentials Without Logarithms." *Labour Economics*, 14(1): 73–98.

Bosquet, Clément, and Hervé Boulhol. 2014. "Applying the GLM Variance Assumption to Overcome the Dcale-dependence of the Negative Binomial QGPML Estimator." *Econometric Reviews*, 33(8): 772–784.

Cameron, A Colin, and Pravin K Trivedi. 2013. *Regression analysis of count data.* 2nd Ed., Cambridge:Cambridge University Press.

Chen, ChinMu, Shih-Yang Lin, Shin-Kun Peng, and Wen-Jen Tsay. 2023. "A Two-Stage Quasi-Maximum Likelihood Method for Poisson Models with Two Endogenous Binary Variables: An Application to Trade and Investment Agreements." Unpublished. Available at SSRN: https://ssrn.com/abstract=4667432.

Chib, Siddhartha, and Edward Greenberg. 1998. "Analysis of Multivariate Probit Models." *Biometrika*, 85(2): 247–361.

Cohn, Jonathan B., Zack Liu, and Malcolm I. Wardlaw. 2022. "Count (and Count-like) Data in Finance." *Journal of Financial Economics*, 146: 529–552.

Deb, Partha, and Pravin K. Trivedi. 2006. "Specification and Simulated Likelihood Estimation of A Non-normal Treatment-outcome Model with Selection: Application to Health Care Utilization." *Econometric Journal*, 6(2): 307–331.

Egger, Peter, Mario Larch, Kevin E. Staub, and Rainer Winkelmann. 2011. "The Trade Effects of Endogenous Preferential Trade Agreement." *American Economic Journal: Economic Policy*, 3(3): 113–143.

Fahlenbrach, Rüdiger. 2009. "Founder-CEOs, Investment Decisions, and Stock Market Performance." *Journal of Financial and Quantitative Analysis*, 44(2): 439–466.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Greene, William H. 2012. *Econometric Analysis.* 7th Ed., Upper Saddle River, NJ:Prentice Hall.

Hausman, Jerry A, Bronwyn H Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, 52(4): 909–938.

Head, Keith, and John Ries. 2008. "FDI as an Outcome of the Market for Corporate Control: Theory and Evidence." *Journal of International Economics*, 74(1): 2–20.

Heckman, James J., and Sergio Urzúa. 2010. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." *Journal of Econometrics*, 156(1): 27–37.

Horn, Roger A., and Fuzhen Zhang. 2005. "Basic Properties of the Schur Complement." In *The Schur Complement and Its Applications.* , ed. Fuzhen Zhang, Chapter 1, 17–46. Springer.

Jochmans, Koen, and Vincenzo Verardi. 2022. "Instrumental-variable Estimation of Exponential-regression Models with Two-way Fixed Effects with an Application to Gravity Equations." *Journal of Applied Econometrics*, 37(6): 1121–1137.

Kenkel, Donald S., and Joseph V. Terza. 2001. "The Effect of Physician Advice on Alcohol Consumption: Count Regression with An Endogenous Treatment Effect." *Journal of Applied Econometrics*, 16(2): 165–184.

Mullahy, John. 1986. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics*, 33(3): 341–365.

Mullahy, John. 1997. "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior." *The Review of Economics and Statistics*, 79(4): 586–593.

Powell, David, and Seth A Seabury. 2018. "Medical Care Spending and Labor Market Outcomes: Evidence from Workers Compensation Reforms." *American Economic Review*, 108(10): 2995–3027.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Santos Silva, J.M.C., and Silvana Tenreyro. 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-maximum Likelihood Estimator." *Economics Letters*, 112(2): 220–222.

Terza, Joseph V. 1998. "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects." *Journal of Economics*, 84(1): 129–154.

Tinbergen, Jan. 1962. *Shaping the World Economy.* New York:Twentieth Century Fund.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50(1): 1–25.

White, Halbert. 1994. *Estimation, Inference and Specification Analysis.* Cambridge University Press.

Windmeijer, Frank A. G., and J. M. C. Santos Silva. 1997. "Endogeneity in Count Data Models: An Application to Demand for Health Care." *Journal of Applied Econometrics*, 12(3): 281–294.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Wooldridge, Jeffrey M. 2014. "Quasi-maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables." *Journal of Econometrics*, 182(1): 226–234.

五、申請補助經費：

（一）請將本計畫申請書之第七項(表CM07)、第八項(表CM08)、第九項(表CM09)、第十項(表CM10)、第十一項(表CM11)、第十二項(表CM12\CM12-1)所列費用個別加總後，分別填入「研究人力費」、「耗材、物品、圖書及雜項費用」、「國外學者來臺費用」、「研究設備費」、「國外差旅費-執行國際合作與移地研究」及「國外差旅費-出席國際學術會議」等欄內。

（二）管理費為申請機構配合執行本計畫所需之費用，其計算方式係依本會規定核給補助管理費之項目費用總和及各申請機構管理費補助比例計算後直接產生，計畫主持人不須填寫「管理費」欄。

（三）依據本會「補助延攬客座科技人才作業要點」規定提出博士級研究人員申請，請依各年度申請之名額填入下表，如於申請時一併提出「補助延攬博士級研究人員員額/人才進用申請書」（表CIF2101、CIF2102），若計畫核定僅核定名額者應於提出合適人選後，另向本會提出進用申請，經審查通過後，始得進用該名博士級研究人員。

（四）申請機構或其他單位（含國內外、大陸地區及港澳）補助項目，請檢附相關證明文件。

金額單位：新臺幣元

執行年次 補助項目		第一年 (113年8月～114年7月)	第二年 (114年8月～115年7月)	第三年	第四年	第五年
業務費		**1,219,454**	**1,220,454**			
研究人力費		1,169,454	1,169,454			
耗材、物品、圖書及雜項費用		50,000	51,000			
國外學者來臺費用		0	0			
研究設備費		**67,000**	**45,000**			
國外差旅費		**155,330**	**146,840**			
執行國際合作與移地研究		0	0			
出席國際學術會議		155,330	146,840			
管理費		**192,968**	**189,818**			
合計		**1,634,752**	**1,602,112**			
博士級研究人員	國內、外地區	共 0 名	共 0 名	共 ____ 名	共 ____ 名	共 ____ 名
	大陸地區	共 0 名	共 0 名	共 ____ 名	共 ____ 名	共 ____ 名

申請機構或其他單位（含國內外、大陸地區及港澳）補助項目（無配合補助項目者免填）

配合單位名稱	配合補助項目	配合補助金額	配合年次	證明文件

六、主要研究人力：

（一）請依照「主持人」、「共同主持人」、「協同研究人員」及「博士級研究人員」等類別之順序分別填寫。

類 別	姓名	服務機構/系所	職稱	在本研究計畫內擔任之具體工作性質、項目及範圍	＊每週平均投入工作時數比率(%)
主持人	蔡文禎	中央研究院經濟研究所	研究員	文獻回顧, 資料收集, 模型設定與求解，模擬以及撰寫報告	30%

※ 註：每週平均投入工作時數比率係填寫每人每週平均投入本計畫工作時數佔其每週全部工作時間之比率，以百分比表示（例如：50%即表示該研究人員每週投入本計畫研究工作之時數佔其每週全部工時之百分五十）。

（二）如依據本會「補助延攬客座科技人才作業要點」規定申請博士級研究人員，請另填表CIF2101及CIF2102(若已有人選者，請務必填註人選姓名，並將其個人資料表(表C301～表C303)併同本計畫書送本會)。

七、研究人力費：

（一）凡執行計畫所需研究人力費用，均得依本會「補助專題研究計畫研究人力約用注意事項」規定，按所屬機構自訂敘薪標準及職銜，就預估專任、兼任人員或臨時工需求填寫，並請述明該研究人力在本計畫內擔任之具體內容、性質、項目及範圍，以利審查。專任人員不限學歷，包含博士級人員。

（二）約用專任人員，請依其於專題研究計畫負責之工作內容，所應具備之專業技能、獨立作業能力、預期績效表現及相關學經歷年資等條件，綜合考量敘薪，並檢附各機構自訂之薪資支給依據，以為本會核定聘用助理經費之參考。

（三）請分年列述。

第 1 年　　金額單位：新臺幣元

類別	金額	請敘明在本計畫內擔任之具體內容、性質、項目及範圍 (如約用專任人員，請簡述其於計畫內所應具備之專業技能、獨立作業能力、預期績效表現及相關學經歷年資等條件)
專任人員	739,386	學士級專任助理(胡毓釗)，始聘於「附加負擔及承諾與垂直結合之限制競爭效果分析」計畫，並以多年競爭法與經濟分析的工作經驗協助蒐集、研析歐美的競爭法案例資料。於此次新申請計畫，負責文獻搜集與資料分析、協助研究員模型應用的工作。 739,386元(含月支費用、年終獎金、勞健保費雇主負擔部分、勞工退休金雇主負擔部分) x 1名
兼任人員(博士候選人-僱傭關係)	430,068	兼任助理(林世揚-博士班六年級學生)。 長年協助研究員處理模型計算及數學推導，並幫忙蒐集資料及論文內容整理。於此次新申請計畫中，將負責推導模型及應用方面的工作。 430,068元(含月支費用、勞健保費雇主負擔部分、勞工退休金雇主負擔部分) x 1名
合計	1,169,454	

第 2 年　　金額單位：新臺幣元

類別	金額	請敘明在本計畫內擔任之具體內容、性質、項目及範圍 (如約用專任人員，請簡述其於計畫內所應具備之專業技能、獨立作業能力、預期績效表現及相關學經歷年資等條件)
專任人員	739,386	學士級專任助理(胡毓釗)，雖始聘於「附加負擔及承諾與垂直結合之限制競爭效果分析」計畫，惟以多年競爭法與經濟分析的工作經驗協助蒐集、研析歐美的競爭法案例資料。於此次新申請計畫，負責文獻搜集與資料分析、協助研究員模型應用的工作。 739,386元(含月支費用、年終獎金、勞健保費雇主負擔部分、勞工退休金雇主負擔部分) x 1名
兼任人員(博士候選人-僱傭關係)	430,068	兼任助理(林世揚-博士班六年級學生)。 長年協助研究員處理模型計算及數學推導，並幫忙蒐集資料及論文內容整理。於此次新申請計畫中，將負責推導模型及應用方面的工作。 430,068元(含月支費用、勞健保費雇主負擔部分、勞工退休金雇主負擔部分) x 1名
合計	1,169,454	

八、耗材、物品、圖書及雜項費用：

(一) 凡執行研究計畫所需之耗材、物品(非屬研究設備者)、圖書及雜項費用，均可填入本表內。
(二) 說明欄請就該項目之規格、用途等相關資料詳細填寫，以利審查。
(三) 若申請單位有配合款，請於備註欄註明。
(四) 請分年列述。

第 1 年　　　　金額單位：新臺幣元

項目名稱	說明	單位	數量	單價	金額	備註
消耗性器材	隨身硬碟	個	1	4,000	4,000	
國內學會之年費	台灣經濟學會、台灣計量經濟學會與中華財政學會等學會的年費	次	3	1,000	3,000	台灣經濟學會年費:1,000元 台灣計量經濟學會年費:1,000元 中華財政學會年費:1,000元
論文發表費	論文投稿費用	次	3	5,000	15,000	
論文潤稿費	論文須經專業潤稿人員進行英文編修，預計往來修正2次	次	2	5,000	10,000	
電腦使用費	購買或申請使用論文相關的資料庫	批	1	5,000	5,000	
郵電費	支付研究計畫相關的郵電費	批	1	1,000	1,000	
雜支	電腦與其周邊設備與印表機等維修費用	次	2	5,000	10,000	
圖書	購買研究計畫相關的圖書	本	1	2,000	2,000	
合計					50,000	

第 2 年　　　　金額單位：新臺幣元

項目名稱	說明	單位	數量	單價	金額	備註
消耗性器材	顯示器螢幕	個	1	5,000	5,000	
國內學會之年費	台灣經濟學會、台灣計量學會與中華財政學會等學會的年費	個	3	1,000	3,000	台灣經濟學會年費:1,000元 台灣計量經濟學會年費:1,000元 中華財政學會年費:1,000元
論文發表費	論文投稿費用	次	3	5,000	15,000	
論文潤稿費	論文須經專業潤稿人員進行英文編修，預計往來修	次	2	5,000	10,000	

	正2次					
電腦使用費	購買或申請使用論文相關的資料庫	批	1	5,000	5,000	
郵電費	支付研究計畫相關的郵電費	批	1	1,000	1,000	
雜支	電腦與其周邊設備與印表機等維修費用	次	2	5,000	10,000	
圖書	購買研究計畫相關的圖書	本	1	2,000	2,000	
合計					51,000	

十、研究設備費：

（一）凡執行研究計畫所需單價在新臺幣一萬元以上且使用年限在二年以上與研究計畫直接有關之各項設備屬之。各類研究設備金額請於金額欄內分別列出小計金額。
（二）購置設備單價在新臺幣二十萬元以上者，須檢附估價單。
（三）若申請機構及其他機構有提供配合款，請務必註明提供配合款之機構及金額。
（四）儀器設備單價超過新臺幣六十萬元(含)以上者，請詳述本項設備之規格與功能(諸如靈敏度、精確度…等)，其他重要特性與重要附件，以及申購本設備對計畫執行之必要性。本項設備若獲補助，主持人應負維護保養之責，並且在不妨礙個人研究計畫或研究群計畫之工作下，同意提供他人共同使用，以避免設備閒置。
（五）計畫主持人執行本項研究計畫，如欲申請購置單價新臺幣壹千萬元(含)以上之大型儀器，請填表CM10-1。該項設備若獲本會核定補助新臺幣壹千萬元(含)以上，則單獨核給一個規劃計畫，主持人須遵守本會大型儀器之管考規定。
（六）經本會補助之大型儀器，儀器資訊須公開於本會全球資訊網之跨部會服務平台「貴重儀器開放共同管理平台」(https://www.nstc.gov.tw/folksonomy/instrument?l=ch)。
（七）請分年列述。

第 1 年 金額單位：新臺幣元

類別	設備名稱（中文/英文）	說明	數量	單價	金額	經費來源	
						本會補助經費需求	提供配合款之機構名稱及金額
儀器及資訊設備	桌上型電腦	購買研究計畫所需的電腦設備 國產27型 i5液晶電腦(AIO A5702/i5-1340P/16G/1T HDD+256G SSD/W11)約為$35,000	1	35,000	35,000	35,000	
儀器及資訊設備	GAUSS 數值分析與統計計量軟體	GAUSS 第23版(學術)(單一使用者執照)/GAUSS v23 (Academic)(Single-user license)價格約為$1,000美金	1	32,000	32,000	32,000	
合計					67,000	67,000	

第 2 年 金額單位：新臺幣元

類別	設備名稱（中文/英文）	說明	數量	單價	金額	經費來源	
						本會補助經費需求	提供配合款之機構名稱及金額
儀器及資訊設備	筆記型電腦	HP EliteBook 640 G10 14吋 商務筆電，內含window作業系統，約為$45,000元	1	45,000	45,000	45,000	
合計					45,000	45,000	

十二、國外差旅費-出席國際學術會議：

（一）計畫主持人及參與研究計畫之相關人員參加國際學術會議得申請本項經費。
（二）請詳述預定參加國際學術會議之性質、預估經費、天數及地點。
（三）機票費、生活費及其他費用之標準，請依照行政院頒布之「國外出差旅費報支要點」規定填列（網址https://law.dgbas.gov.tw/LawContent.aspx?id=FL017584）。
（四）請詳述計畫主持人近三年參加國外舉辦之國際學術會議論文之發表情形。（包括會議名稱、時間、地點、發表之論文題目、補助機構，及後續收錄於期刊或專書之名稱、卷號、頁數、出版日期）
（五）請分年列述。

第 1 年　　金額單位：新臺幣元

出席國際學術會議			
出席國際學術會議人數	共 1 名	金　額	155,330
費用說明	參加 Western Economic Association International 舉辦的WEAI第100屆年會，此會議匯集太平洋兩岸的頂尖經濟學家於一堂，分享最新的研究及方法論。 1.會議時間：預計2025年6月於美國舊金山舉行，從6月20日到6月24日，為期五天。 2.預估經費：包含：機票、住宿與相關費用，總計為155,330元 機票：NTD.62,000元 住宿生活費：USD.440*6(晚)*32(匯率)=NTD.84,480元 註冊費、保險費等其他費用 ：NTD.8,850元		
近三年論文發表情形	1. 會議名稱：6th International Conference on Econometrics and Statistics 會議時間：2023.08.01~03 會議地點：日本東京早稻田大學 論文題目：Consistent Autoregressive Spectral Estimates under GARCH-type Noises 補助機構：中央研究院 收錄期刊：working paper		

第 2 年　　金額單位：新臺幣元

出席國際學術會議			
出席國際學術會議人數	共 1 名	金　額	146,840
費用說明	參加 Western Economic Association International 舉辦的WEAI第101屆年會，此會議匯集太平洋兩岸的頂尖經濟學家於一堂，分享最新的研究及方法論。 1.會議時間：2026年6月於美國科羅拉多州丹佛市舉行，從6月29日至7月3日，為期五天。 2.預估經費：包含：機票、住宿與相關費用，總計為146,840元 機票：NTD.75,000元 住宿生活費：USD.320*6(晚)*32(匯率)=NTD.61,440元 註冊費、保險費等其他費用 ：NTD.10,400元		
近三年論文發表情形	1. 會議名稱：6th International Conference on Econometrics and Statistics 會議時間：2023.08.01~03 會議地點：日本東京早稻田大學		

	論文題目：Consistent Autoregressive Spectral Estimates under GARCH-type Noises 補助機構：中央研究院 收錄期刊：working paper

十四、近三年內執行本會之所有計畫

計畫名稱 （本會補助者請註明編號）	計畫內擔任之工作	起迄年月	補助或委託機構	執行情形	經費總額
附加負擔及承諾與垂直結合之限制競爭效果分析(112-2410-H-001-090-)	主持人	2023/08/01～2024/07/31	國家科學及技術委員會	執行中	937,000
運用臺灣普查資料檢驗競爭政策低度執法風險之實証分析(111-2410-H-001-067-)	主持人	2022/08/01～2023/07/31	國家科學及技術委員會	已結案	888,000
市場集中度與競爭法執法及資訊科技發展(110-2410-H-001-084-)	主持人	2021/08/01～2022/10/31	國家科學及技術委員會	已結案	772,000
運用主計處工商普查檢驗市場集中度與企業利潤的關係(109-2410-H-001-042-)	主持人	2020/08/01～2021/07/31	國家科學及技術委員會	已結案	761,000
合　　計					3,358,000

國家科學及技術委員會人文處專題研究計畫

主持人代表性研究成果表(2023.10.25 製表)

計畫主持人：__________________________ **職稱：**_____________
服 務 機 關 ：__________________________

＊ 本表為評量計畫主持人研究表現之依據，務請詳實填寫。如非所上傳之代表性研究成果，則不列入評量。
＊ 頁數以 5 頁為限（字型大小 12，標準字元間距，單行間距）。

計畫主持人若於申請截止日前 10 年內曾生產、請育嬰假或曾服國民義務役者，應檢附相關證明文件，並請填寫代表性研究成果之期限得依規定延長____年[註1]。

一、請列出所上傳近 10 年(2014.01~2023.12)代表性研究成果（可含實作成果）至多 5 篇(本)，其中至少 1 篇(本)為近 5 年之研究成果 [註2]。

1. Liao, J.C. and Tsay, W.J. 2020. “Optimal Multi-step VAR Forecast Averaging.” *Econometric Theory*, 36(6), 1099-1126.
2. Chu, C.Y.C., Lin, J.C. and Tsay, W.J. 2020. “Males' Housing Wealth and Their Marriage Market Advantage.” *Journal of Population Economics*, 33, 1005-1023.
3. Hwu, S.T., Fu, T.T. and Tsay, W.J. 2021. “Estimation and Efficiency Evaluation of Stochastic Frontier Models with Interval Dependent Variables.” *Journal of Productivity Analysis*, 56, 33-44.
4. Lai, H. P., and Tsay, W.J. 2018. “Maximum Simulated Likelihood Estimation of the Panel Sample Selection Mode.” *Econometric Reviews*, 37(7), 744 -759.
5. Chu, C.Y.C., Kim, S., and Tsay, W.J. 2014. “Coresidence with Husband’s Parents, Labor Supply, and Duration to First Birth.” *Demography*, 51(1), 185-204.

二、請簡述上述代表性研究成果之創見及對學術、實務或社會之重要貢獻。

1. 本人與廖仁哲教授合作探討 Hansen (2007) 所提的 Mallows model averaging 方法在多變量時間數列的情境下如何修改其目標公式，並證明出在重複一期預測與多期預測下，我們所提出的方法仍保有Hansen (2007)所提的極限最佳性 (asymptotic optimality property)。該文章是使用 Mallows model averaging 方法在多變量時間數

[註1] 計畫主持人於申請截止日前 10 年內曾生產、請育嬰假者，其代表性研究成果之期限，包含其中至少 1 篇(本)為近 5 年之研究成果，得依每一出生數再延長 2 年；曾服國民義務役者，得依實際服役時間予以延長。前述情形應檢附相關證明文件。
[註2] 如為代表性研究成果之期限內被接受刊登尚未出版之著作，應檢附被接受證明文件。

列模型的首例，在文獻與方法的發展上有相當的突破。而且我們的方法是建構在多變量時間數列文獻上最常被使用的 vector autoregressive (VAR)模型，預期將有廣大的應用可能性。

2. 本人與朱敬一院士及他的學生林柔均研究年輕男人是否會因握有更多的財產而使結婚的機會提高，這個看來直覺的問題，文獻上從來沒有能以個體資料進行檢定。本計畫因得到財政部的稅務資料成功將個人財產與其父母的財產整合，以處理年輕人的財產很有可能是由父母贈與而產生的資料整合與可能的內生估計問題。另外，不動產與動產的分配也是另一個可能的內生問題。為此一新議題，我們運用臺灣最近一次的遺贈稅改革作為外生政策變數以處理此風險模型下的多重內生問題。本研究不僅有方法上的創新，實證上也發現持有不動產才能顯著促使年輕男人的結婚機率上升，同時也間接證明不動產為地位財的假說，本文是文獻上第一篇運用個體資料達到此一結果的文章。

3. 本人與東吳傅祖壇教授與美國加州州立大學的胡世唐教授合作探討隨機邊界模型的被解釋變數為區間而非連續變數時的估計與效率計算問題。換言之，我們將一般的 continuous dependent variable 隨機生產邊界模型 (該模型為生產效率分析最常用的模型)推廣至 interval dependent variable 的情境，並依生產效率與成本效率模型的不同，分別算出對應的複合誤差項的累積機率分配函數以進行之後的概似估計法的必要成分。而且我們也導出伴隨區間生產與成本效率模型必須評價的效率公式，該公式是解析解，不需依賴任何複雜的模擬抽樣方式。因此，我們的研究不僅成功擴展隨機邊界模型的適用領域，由連續性的被解釋變數推升至區間數據，為五十多年的重大突破，我們的方法的簡便性，也將使應用範圍大增。

4. 本人與中正大學的賴宏彬教授以新的視角重新討論 panel sample selection 的最大概似估計問題。該模型非常受學界重視，原因在於其廣大的應用範圍，美中不足的是其在估計上有著相當的難度，其中包含多重積分的計算問題。基於此，本文利用 closed skew normal 的機率性質重新導出更為清晰和方便計算的概似值計算法，並將整個概似估計化簡為單一積分形式，不僅理論上的精確度大為提升，而且模擬實驗也證明出有相當傑出的小樣本表現，相信能為未來的相關實證研究提出更方便可靠的方法。

5. 本人與本所朱敬一院士及當時在西雅圖華盛頓大學的Seik Kim 教授探討結婚時是否與男方父母共住對之後生育第一胎的期間的影響為何。我們知道女性的生育決定受其勞動參與以及上述共住決定的影響。我們是第一篇文章點出該兩項決策應是內生決定變數，並據此擴充傳統的Duration 模型使其在endogenous dummy variables 存在的情形下仍能進行估計與檢定。更重要的是我們發現共住會拖延生育第一胎的時程，而非文獻上數以百計所得到會使生育變快的結果。這在直觀上相當合理，因為共住本身即代表隱私權的不完整，加上與公婆共住自有各方面的壓力產生。本文因此可能改變人們對於家庭結構與生育決定之間互動的看法，本文已發表在人口學的

旗艦期刊 Demography。