Portfolio Diversification and Complementarity in Asset Demand Systems

Ozan E. Akbas & Ao Wang

December 2024 No: 1533

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Portfolio Diversification and Complementarity in Asset Demand Systems [*]

Ozan E. Akbas [†] **Ao Wang** [‡]

This Version: 15 December 2024
First Version: 15 December 2024

Abstract

Investors evaluate their entire portfolio, not individual assets, striving to balance returns and risks through effective diversification. This paper introduces a flexible demand system accommodating heterogeneous substitution, cross-asset complementarities, and diverse investment strategies. By relaxing multinomial logit assumptions, our model better captures portfolio allocation decisions, linking portfolio weights to both individual asset and portfolio-wide characteristics. We propose a demand-inverse approach for the identification of structural parameters. This approach implies a Generalized Method of Moments estimation procedure with novel instruments addressing cross-asset dependencies. Monte Carlo simulations validate the model, demonstrating improved finite-sample properties over standard multinomial logit frameworks.

JEL Codes: C51, G11, G23
Keywords: asset demand systems, flexible substitution, cross-asset complementarity

---

[*] We are thankful for useful comments by Andrea Gamba, Ruggero Jappelli and seminar participants at Warwick Business School Finance Group Brown Bag Seminar.

[†] Warwick Business School, University of Warwick. E-mail: ozan.akbas@warwick.ac.uk
[‡] Department of Economics, University of Warwick and CAGE Research Centre. E-mail: ao.wang@warwick.ac.uk

1. Introduction

The conventional wisdom in portfolio choice is that investors consider their entire portfolio rather than evaluating individual assets in isolation. In practice, this process is rather messy: not all assets are close substitutes, and some may even complement one another when combined. For example, a hedge fund might balance high-growth tech stocks against stable blue-chip firms, exploiting their different characteristics and correlations to manage overall risk more effectively. Yet much of the demand-systems asset pricing literature treats stocks as independent substitutes, overlooking the critical role these cross-asset dependencies play in strategic asset allocation. This simplification ignores how investors rely on diversification, nonlinear interactions, and nuanced risk-return trade-offs.

In this paper, we develop a new characteristic-based demand model that relaxes these restrictive assumptions and propose an estimation strategy that can still leverage mandate-based instruments. Our model captures heterogeneous substitution patterns and allows for asset complementarity, offering a more realistic depiction of how characteristics shape expected returns and risk exposures. In particular, it enables characteristics to increase expected returns without necessarily an one-for-one increase in systematic risk. In our framework, an asset's portfolio weight depends on its own characteristics as well as those of other assets in the same portfolio. Consequently, when asset pairs jointly reduce overall variance, hence enhancing diversification, they are complements.

Furthermore, we propose a demand-inverse approach to identify our model's structural parameters and adopt a Generalized Method of Moments (GMM) estimation procedure. In addition to mandate-based price instruments, we leverage other assets' characteristics in the same portfolio as additional instruments accounting for cross-asset dependencies. Through Monte Carlo simulations, we show that ignoring cross-asset dependencies introduces significant bias in price elasticities. Our model significantly reduces this bias, offering substantial improvements over the commonly used approach of estimating asset demand proposed by Koijen and Yogo (2019) (henceforth KY19).

The multinomial logit (MNL) demand structure in KY19 implicitly assumes that assets are good substitutes via the Independence of Irrelevant Alternatives (IIA) property. The IIA implies that the relative demand for any two asset is unaffected by the characteristics of other assets. This simplification overlooks a critical aspect of mean-variance optimal portfolio choice (Markowitz, 1952): portfolio weights are determined not only by expected returns and variances but also by covariances between assets. By assuming independence between assets, the MNL framework fails to capture the role of diversification, and overall how cross-asset dependencies influence portfolio allocation.

Another implication of the IIA in KY19 is that the one-factor structure forces any increase in returns to be matched by a uniform increase in factor loadings, leaving no room for alpha. In contrast, if omitted risk factors or informational inefficiencies allow for characteristic-driven return differentials not fully captured by the factor loadings, alpha can emerge. Such scenarios are arguably more representative of complex institutional portfolios, where systematic risk and pricing inefficiencies interact in nuanced ways.

To fix the ideas, we gradually build on an example throughout this paper:

Example. *Consider two investors representing two pension funds with investment mandates on the allocation of their assets under management. Alice's (Bob's) mandate requires a fixed weight allocation to value stocks (small-cap stocks). These mandates reflect different investment strategies: Alice's strategy focuses on investing in undervalued firms with solid fundamentals, while Bob's strategy targets firms with smaller market capitalization that may offer* *higher growth potential. Ideally, a demand model should capture these differences in how characteristics like book-to-market ratio (as a proxy for value) or market equity (as a proxy for size) translate into variations in risk-return relationship. However, using the linear MNL framework of KY19 forces increases in expected returns to be matched by proportionate increases in factor loadings; same proportional risk-return trade-off for the value strategy as for the size strategy. Thus, the model leaves no room for returns that exceed what the single factor can explain.*

Focusing on the Capital Asset Pricing Model (CAPM) as the benchmark most often used by academics and practitioners, we see that investors like Alice and Bob indeed achieve abnormal returns relative to what CAPM predicts. The results are summarized in Table 1 for quarterly U.S. institutional equity holdings based on SEC 13F filings via FactSet, combined with quarterly U.S. equity data from the merged Compustat-CRSP dataset, spanning 2000.Q1 to 2022.Q4. First, we aggregate stock returns and characteristics within each investor portfolio into portfolio-level returns and characteristics using portfolio weights. Next, we sort investors into deciles based on one-year lagged characteristics. Finally, we compute equally-weighted average returns and CAPM alphas.

Non-zero alphas across all sorting variables indicate that the CAPM fails to fully account for systematic risks tied to these characteristics, aligning with well-documented anomalies in asset pricing (e.g., most recently in Lopez-Lira and Roussanov (2023)). For instance, portfolios sorted on market equity exhibit a pronounced size effect, with small-cap focused investors in the 1^{st} decile achieving substantially higher returns and alphas relative to small-cap focused investors in the 10^{th} decile. These estimates highlight the limitations of KY19's linear MNL framework in capturing investors' heterogeneous sensitivity to asset characteristics and their alpha-seeking behavior.

Table 1: Anomalies in Institutional Equity Portfolios

This table reports anomalies in investor equity portfolio returns along dimensions of asset characteristics considered in most demand systems asset pricing frameworks: market equity, book-to-market ratio, investment (log growth of assets), and profitability (operating profit to book equity). First, we aggregate stock returns and characteristics within each investor portfolio into portfolio-level returns and characteristics using portfolio weights. Next, we sort investors into deciles based on one-year lagged characteristics. Finally, we compute equally-weighted average returns and CAPM alphas (percent). Our sample includes quarterly U.S. institutional equity holdings based on SEC 13F filings via FactSet, combined with quarterly U.S. equity data from the merged Compustat-CRSP dataset, spanning 2000.Q1 to 2022.Q4.

Deciles	1	2	3	4	5	6	7	8	9	10
Panel A: Market Equity										
Excess Returns	6.73	5.39	4.81	4.38	4.03	3.77	3.61	3.48	3.38	3.34
CAPM Alpha	4.34	3.31	2.87	2.51	2.23	2.01	1.87	1.77	1.67	1.60
Panel B: Book to Market										
Excess Returns	5.42	4.39	4.00	3.95	3.91	3.95	4.07	4.28	4.39	4.56
CAPM Alpha	3.33	2.55	2.21	2.19	2.15	2.18	2.27	2.42	2.46	2.45
Panel C: Investment										
Excess Returns	4.88	3.94	3.79	3.79	3.77	3.86	4.02	4.29	4.76	5.82
CAPM Alpha	2.95	2.20	2.06	2.05	2.00	2.08	2.19	2.40	2.75	3.50
Panel D: Profitability										
Excess Returns	6.48	5.20	4.66	4.26	3.95	3.82	3.68	3.60	3.58	3.71
CAPM Alpha	4.02	3.09	2.71	2.41	2.16	2.06	1.96	1.90	1.88	1.99

Related Literature. This paper relates to two strands of research on discrete choice models of demand: (i) the finance literature on demand systems in asset pricing and (ii) the empirical industrial organization (IO) literature on demand with complementarity. The finance literature, inspired by KY19, has predominantly relied on the MNL framework. While computationally convenient, this approach overlooks heterogeneous substitution patterns arising from investors' distinct strategies. Our paper extends this literature by introducing a more flexible model that incorporates heterogeneous risk-return profiles within investor portfolios, thereby allowing for richer patterns of substitutability and complementarity.

KY19 started the literature on estimating asset demand systems. This methodological approach has been used to study the crucial impact of various investor's demand on prices in different asset classes, ranging from equities (Koijen and Yogo, 2019; van der Beck and Jaunin, 2023; Huebner, 2023; Mainardi, 2023; Gabaix et al., 2023; Noh et al., 2023; Koijen et al., 2023; Haddad et al., Forthcoming; Gabaix and Koijen, 2024), corporate bonds (Koijen and Yogo, 2023; Chaudhary et al., 2023; Fang, 2024; Darmouni et al., 2024; Siani, 2024), government bonds (Koijen et al., 2021; Jansen et al., 2024; Eren et al., 2024), and currencies (Koijen and Yogo, 2024; Jiang et al., 2024, Forthcoming). Our work builds on this literature by addressing biases introduced by the IIA assumption and by allowing for more flexible substitutability and complementarity.

Recent studies have sought to relax the linear MNL framework. For instance, Chaudhary et al. (2023) use a nested logit structure to model flexible substitution across credit ratings in corporate bond markets. Similarly, Allen et al. (2024) adopt a setup akin to an Almost Ideal Demand System (Deaton and Muellbauer, 1980) to document varying substitution patterns across Canadian Treasury bond maturities. On the theoretical side, Fuchs et al. (2024) highlight that portfolio choice inherently features cross-asset demand complementarities driven by diversification motives, raising concerns about the validity of elasticity estimates derived from demand systems. Davis et al. (2024) further argue that high substitutability assumptions in classical asset pricing models are a primary driver of inelastic demand estimates for mean-variance investors. Our paper differs from the recent works and advances this literature by addressing the bias introduced by the high substitutability assumption underlying the linear MNL framework. We propose a structural model that incorporates heterogeneous substitution patterns via investors' flexible risk-return trade-offs, offering a framework researchers can apply to estimate elasticities across any asset class.

Finally, our paper bridges the finance and empirical IO literatures by incorporating potential complementarity into asset demand systems. Complementarity can arise in differentiated product markets that feature multiple purchases, such as grocery (Deaton and Muellbauer, 1980; Dubé, 2004; Thomassen et al., 2017; Ershov et al., Forthcoming; Fosgerau et al., 2024; Iaria and Wang, 2020, 2024; Wang, 2024), newspapers (Gentzkow, 2007; Fan, 2013), telecommunication (Liu et al., 2010; Grzybowski and Verboven, 2016), and media (Crawford and Yurukoglu, 2012; Crawford et al., 2018). Our asset demand model, a flexible MNL framework, is microfounded by the mean-variance optimal portfolio choice. The complementarity has a different nature from that in IO demand models (e.g., consumption synergy, preference for variety, shopping cost). Methodologically, our demand-inverse approach resembles the method pioneered by Berry (1994); Berry et al. (1995) (aka BLP) and the recent one in Wang (2024) that incorporates complementarity in the BLP framework. Both ours and theirs consist of a first step of demand inverse and a second step that instruments out unobserved demand shocks, implying a GMM estimation procedure. Because of distinct microfoundations, the invertibility argument in our first step and the instrument validity in the second step differ from those in the BLP approach.

2. Model

We develop a structural asset pricing model from the mean-variance optimal portfolio choice of heterogeneous long-only investors. The optimal portfolio varies across investors due to the heterogeneous nature of the portfolio asset composition. Following KY19, we assume that the asset returns have a one-factor structure, and the expected returns and the factor loadings on this one factor depend on asset characteristics. We show that this framework results in a characteristic-based asset demand system where the portfolio weights of each asset depends on the characteristics of all assets in the investor's portfolio.

2.1. Notation and Setup

Throughout the remainder of this paper, let the indices $i = 1, \ldots, I$, $n = 1, \ldots, N$, and t represent investors, assets, and time, respectively. The investment universe of investor i at time t is denoted by $\mathcal{N}_{i,t} \subseteq \{1, \ldots, N\}$, and is determined by the investor's investment mandate. We refer to the $|\mathcal{N}_{i,t}|$ assets in investor i's portfolio at time t as *inside assets*; all wealth outside of the inside assets are referred to as *outside asset*. We use lowercase letters to denote the natural logarithms of corresponding uppercase variables. For example, let $q_{i,t}(n) = \ln(\, Q_{i,t}(n)\,)$ represent the natural logarithm of investor i's holdings of asset n at time t. The corresponding vector forms are denoted in bold, such that $\mathbf{q}_{i,t} = \ln(\, \mathbf{Q}_{i,t}\,)$.

2.2. Mean-Variance Portfolio Choice for Heterogeneous Investors

Investor i at time t allocates her wealth $A_{i,t}$ across inside assets and an outside asset. Each inside asset n has a gross return $R_{t+1}(n)$ from time t to time $t+1$. The outside asset has a gross return $R_{t+1}(0)$ from time t to time $t+1$. The investor puts $\mathbf{w}_{i,t}$ of her wealth into the inside assets, resulting in portfolio returns $\mathbf{R}^p_{t+1}$:

$$\mathbf{R}^p_{t+1} = \mathbf{w}'_{i,t}\mathbf{R}_{t+1} + (1 - \mathbf{w}'_{i,t}\mathbf{1})R_{t+1}(0) = R_{t+1}(0) + \mathbf{w}'_{i,t}\left(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1}\right), \tag{1}$$

The resulting wealth is then governed by the following intertemporal budget constraint:

$$A_{i,t+1} = A_{i,t}\left(R_{t+1}(0) + \mathbf{w}'_{i,t}\left(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1}\right)\right). \tag{2}$$

Furthermore, the portfolio weights are subject to the intratemporal budget constraint that their sum cannot exceed one, allowing for a positive allocation to the outside asset:

$$\mathbf{1}'\mathbf{w}_{i,t} \leq 1. \tag{3}$$

The $|\mathcal{N}_{i,t}|$-dimensional vector of asset weights in her portfolio $\mathbf{w}_{i,t} \in [0,1]$ maximizes the investor's expected log utility over her terminal wealth at time T:

$$\max_{\mathbf{w}_{i,t}} \ \mathbb{E}_{i,t}\left(\ln\left(A_{i,T}\right)\right) \quad \text{such that} \quad \begin{cases} A_{i,t+1} = A_{i,t}\left(R_{t+1}(0) + \mathbf{w}'_{i,t}\left(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1}\right)\right), \\ \mathbf{1}'\mathbf{w}_{i,t} \leq 1. \end{cases} \tag{4}$$

Lemma 1 (Solution to the Mean-Variance Portfolio Choice Problem). *The optimal* *portfolio weights satisfy*

$$\mathbf{w}_{i,t} = \Sigma_{i,t}^{-1} \left(\mu_{i,t} - \lambda_{i,t} \mathbf{1} \right), \tag{5}$$

where $\lambda_{i,t}$ is the Lagrange multiplier on the intratemporal budget constraint (3).

Proof of Lemma 1. *See Appendix A.1.* □

2.3. Characteristic-Based Demand and Cross-Asset Dependencies

Investor i at time t forms expectations about an asset n's returns based on its observed and unobserved characteristics. The observed characteristics are log market equity (me_t), and other five characteristics $\mathbf{x}_{k,t}(n)$ indexed by k including log book equity, dividends to book equity, operating profits to book equity, log growth of assets, and market beta. Any further characteristic that investor i considers but is unobserved by the econometrician is denoted by $\ln(\epsilon_{i,t}(n))$. Let $\mathbf{x}_{i,t}(n)$ denote the stacked vector of characteristics and $\mathbf{y}_{i,t}(n)$ their Mth-order polynomial for the assets in investor i's portfolio at time t:

$$\mathbf{x}_{i,t}(n) = \begin{bmatrix} \mathrm{me}_t(n) \\ \mathbf{x}_{k,t}(n) \\ \ln(\epsilon_{i,t}(n)) \end{bmatrix} \quad \text{and} \quad \mathbf{y}_{i,t}(n) = \begin{bmatrix} \mathbf{x}_{i,t}(n) \\ \mathrm{vec}\left(\mathbf{x}_{i,t}(n)\mathbf{x}_{i,t}(n)'\right) \\ \vdots \end{bmatrix}. \tag{6}$$

Note that these asset-level characteristics directly link to the market level five-factor model of Fama and French (2015), e.g. via market beta (MKT), market equity (SMB), book-to-market equity (HML), profitability (RMW), investment (CMA). Following KY19, we assume that returns have a one factor structure at the market level, and that expected excess returns and factor loadings on this one factor are affine in asset characteristics.

Assumption 1 (Factor Structure of Returns and Characteristic Dependence). *Asset* *returns exhibit a one-factor structure:*

$$\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1} = a_{i,t} + \Gamma_{i,t}\left(\mathbf{R}^M_{t+1} - R_{t+1}(0)\mathbf{1}\right) + \ \eta_{i,t}, \tag{7}$$

$$\mu_{i,t}(n) = \alpha_{i,t} + \Gamma_{i,t}\,\mathbb{E}_{i,t}\left(\mathbf{R}^M_{t+1} - R_{t+1}(0)\mathbf{1}\right), \tag{8}$$

where $\mu_{i,t}(n) = \mathbb{E}_{i,t}\left(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1}\right)$ *is the expected excess returns relative the outside option,* $Var\left(\eta_{i,t}\right) = \gamma_{i,t}\mathbf{I}$ *is the idiosyncratic variance, and the covariance matrix of returns:*

$$\Sigma_{i,t} = \Gamma_{i,t}\Gamma'_{i,t} + \gamma_{i,t}\mathbf{I}. \tag{9}$$

The expected excess returns and factor loadings are affine in asset characteristics

$$\mu_{i,t}(n) = \phi_{i,t} + \mathbf{y}_{i,t}(n)'\Phi_{i,t}, \tag{10}$$

$$\Gamma_{i,t}(n) = \psi_{i,t} + \mathbf{y}_{i,t}(n)'\Psi_{i,t}, \tag{11}$$

where $\Phi_{i,t}$ *and* $\Psi_{i,t}$ *are vectors and* $\phi_{i,t}$ *and* $\psi_{i,t}$ *are scalars that are constant across assets.*

The investor cares about the trade-off between risk - via the covariance matrix $\Sigma_{i,t}$ - and return - via the expected excess returns $\mu_{i,t}(n)$. Both of these components are tied to the characteristics of the assets in the portfolio via Assumption 1. Building on this parametrization, the mean-variance optimal portfolio choice simplifies to a characteristics-based asset demand system, where the optimal portfolio weights $w_{i,t}(n)$ are affine in characteristics of the assets in the investor's portfolio.

Lemma 2 (Characteristics-Based Asset Demand). *The optimal portfolio weights are affine* *in asset characteristics:*

$$w_{i,t}(n) = \pi_{i,t} + \mathbf{y}'_{i,t}(n)\mathbf{\Pi}_{i,t}, \tag{12}$$

where $\mathbf{\Pi}_{i,t}$ and $\pi_{i,t}$ capture how characteristics influence portfolio weights each investor:

$$\mathbf{\Pi}_{i,t} = \frac{1}{\gamma_{i,t}}(\mathbf{\Phi}_{i,t} - \mathbf{\Psi}_{i,t}\kappa_{i,t}), \tag{13}$$

$$\pi_{i,t} = \frac{1}{\gamma_{i,t}}(\phi_{i,t} - \lambda_{i,t} - \psi_{i,t}\kappa_{i,t}). \tag{14}$$

The term $\kappa_{i,t}$ reflects the risk-adjusted expected excess returns contributed by the single factor:

$$\kappa_{i,t} = \frac{\Gamma'_{i,t}\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1}\right)}{\Gamma'_{i,t}\Gamma_{i,t} + \gamma_{i,t}}. \tag{15}$$

Proof of Lemma 2. *See Appendix A.2.* □

Ultimately, $\kappa_{i,t}$ is a portfolio-wide variable that depends on the characteristics of all assets in the portfolio. When the characteristics of an asset m in a portfolio changes, then, all else equal, the risk exposure of other $-m$ assets in the same portfolio are adjusted by the changing portfolio risk exposure. Thus, without further restrictions as in KY19, the own characteristics of an asset is not the only source of variation in its portfolio weights.

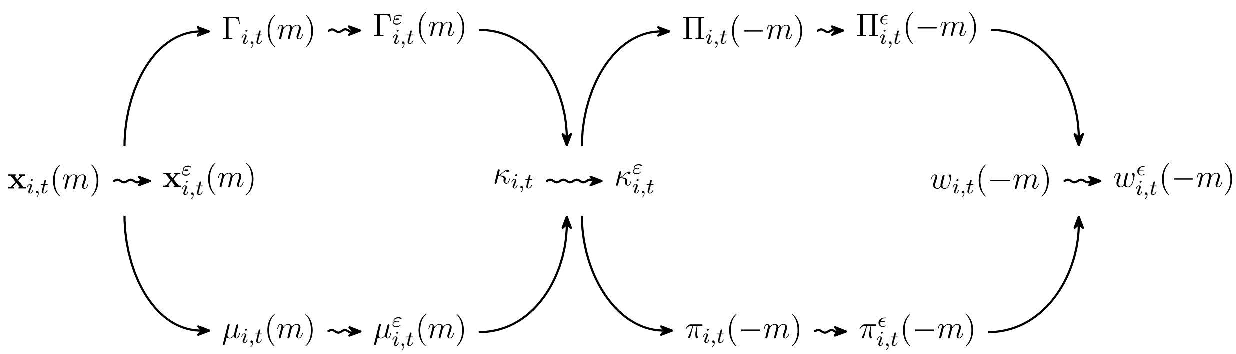

Figure 1: Cross-Asset Dependency

This figure shows the cross-asset dependency in the characteristic-based asset demand derived from mean-variance portfolio choice. In a portfolio, when the asset m's characteristics change by an amount ε, i.e, $\mathbf{x}^{\varepsilon}_{i,t}(m) = \mathbf{x}_{i,t}(m) + \varepsilon$, then, all else equal, the portfolio weights of all other $-m$ assets change.

To model institutional portfolios realistically, it is important to account for wealth allocated outside the set of modeled assets. The outside asset represents any unobserved portion of the investor's wealth, e.g., if the researcher is modeling institutional investor's equity demand, then the outside asset includes any equities with missing information in the dataset, fixed-income securities, or any other asset classes not considered in estimation. By formalizing the role of this outside asset, we ensure that the modeled portion of the portfolio fits into the investor's broader wealth allocation, maintaining internal consistency and economic credibility.

Assumption 2 (Parametrizing the Portfolio Weight on the Outside Asset). *The proportion of wealth allocated to the outside asset $\pi_{i,t}$ is equal to the portfolio weight assigned to the outside asset $w_{i,t}(0)$:*

$$\pi_{i,t} = w_{i,t}(0). \tag{16}$$

Linking the outside asset's share of wealth directly to $w_{i,t}(0)$ serves two purposes. First, it aligns the theoretical representation of unmodeled wealth with the empirical measure, because any fraction of wealth not allocated to the modeled assets is treated as an outside option, simplifying both interpretation and implementation. Second, it allows for a clean, empirically tractable partition between modeled and unmodeled assets.

Moreover, realistic portfolios rarely comprise only the modeled subset of assets, e.g., a large institutional investor only investing in equities. Investors likely keep some residual wealth in other forms, reflecting liquidity preferences, regulatory constraints, or simply unobserved investment options. Ensuring a positive outside asset weight mimics these conditions and preserves the model's applicability to actual investment scenarios.

Assumption 3 (Positive Weight on the Outside Asset). *We assume that the Lagrange multiplier associated with the intratemporal budget constraint Equation 3 is zero:*

$$\lambda_{i,t} = 0. \tag{17}$$

This assumption ensures that the portfolio remains realistic and avoids corner solutions where all wealth is allocated to the modeled assets. By allowing a residual allocation to the outside asset, we prevent the portfolio from becoming artificially constrained. Hence, our model reflects real-world investor behavior where portfolios extend beyond a narrow set of securities.

Having established the role of the outside asset, the next challenge is identifying the structural parameters of our characteristic-based model. Note that under Lemma 2 and

Assumptions 2, 3, we can express the portfolio weights (12) in relative terms:

$$\frac{w_{it}(n)}{w_{it}(0)} = 1 + \mathbf{y}'_{i,t}(n)\frac{\mathbf{\Pi}_{it}}{\pi_{it}} = 1 + \mathbf{y}'_{i,t}(n)\left(\frac{\frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}} - \frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}\kappa\left((y_{i,t}(n))_{n\in\mathcal{N}_{i,t}}; \frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}}, \frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}, \frac{\gamma_{i,t}}{\psi^2_{i,t}}\right)}{1 - \frac{\psi_{i,t}}{\phi_{i,t}}\kappa\left((y_{i,t}(n))_{n\in\mathcal{N}_{i,t}}; \frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}}, \frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}, \frac{\gamma_{i,t}}{\psi^2_{i,t}}\right)}\right),$$

where

$$\kappa\left((y_{i,t}(n))_{n\in\mathcal{N}_{i,t}}; \frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}}, \frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}, \frac{\gamma_{i,t}}{\psi^2_{i,t}}\right) = \frac{\sum_{n\in\mathcal{N}_{i,t}}\left(1 + y'_{i,t}(n)\frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}\right)\left(1 + y'_{i,t}(n)\frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}}\right)}{\sum_{n\in\mathcal{N}_{i,t}}\left(1 + y'_{i,t}(n)\frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}\right)^2 + \frac{\gamma_{i,t}}{\psi^2_{i,t}}}.$$

In this expression for portfolio weights relative to the outside asset, the constant terms $(\phi_{i,t}, \psi_{i,t})$ and the coefficient vectors $(\mathbf{\Phi}_{i,t}, \mathbf{\Psi}_{i,t}, \gamma_{i,t})$ introduce a scale ambiguity. Without a normalization, only the ratios $\frac{\mathbf{\Phi}_{i,t}}{\phi_{i,t}}$, $\frac{\mathbf{\Psi}_{i,t}}{\psi_{i,t}}$, and $\frac{\gamma_{i,t}}{\psi^2_{i,t}}$ are identifiable, as scaling both numerator and denominator by the same factor leaves these ratios unchanged while altering the absolute levels of $\phi_{i,t}$, $\psi_{i,t}$, $\mathbf{\Phi}_{i,t}$, $\mathbf{\Psi}_{i,t}$, and $\gamma_{i,t}$. This rescaling preserves model predictions but renders the absolute interpretation of these parameters ambiguous. Consequently, the model cannot distinguish between genuine shifts in returns and factor loadings versus arbitrary rescaling of parameter values, necessitating a normalization to resolve this ambiguity.

Assumption 4 (Identification of Relative Differences). *Without loss of generality, we set the constants in* (10) *and* (11) *to unity:*

$$\phi = \psi = 1. \tag{18}$$

By fixing ϕ and ψ at unity, we remove arbitrary scaling and anchor the model's reference point for expected returns and factor loadings. Hence, the relative differences, i.e., how characteristics shift returns and factor exposures, become identifiable. If we had chosen different constants, e.g., $\phi = 2$ and $\psi = 1$, the structure of the model's predictions would remain unchanged, confirming that this is a normalization rather than a substantive economic restriction. Ultimately, this step ensures a coherent estimation process such that the effects of characteristics on risk and return can be estimated without confounding scale ambiguities, enabling meaningful inference about investor demand and portfolio formation.

Building on the normalization Assumption 4, we define linear transformations of asset characteristics that capture the expected returns and factor loadings, respectively. These definitions operationalize the relationship between portfolio weights and asset characteristics, enabling a systematic characterization of investor preferences:

$$U_{i,t}(n) = U_{i,t}(\mathbf{x}_{i,t}(n)) = 1 + \mathbf{y}'_{i,t}(n)\Phi_{i,t}, \tag{19}$$

$$V_{i,t}(n) = V_{i,t}(\mathbf{x}_{i,t}(n)) = 1 + \mathbf{y}'_{i,t}(n)\Psi_{i,t}. \tag{20}$$

Together with the feasibility assumptions (2) and (3), we get a structural model of characteristic-based asset demand that incorporates portfolio-wide risk considerations:

Proposition 1 (Cross-Asset Dependency in Portfolio Weights). *A mean-variance investor's demand for an asset in her investment universe, relative to the outside asset depends on three key factors: (i) the asset's expected returns relative to the outside asset's returns* $U(\cdot)$*,* *(ii) its factor loading $V(\cdot)$, and (iii) portfolio-wide systematic risk exposure κ:*

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = 1 + \mathbf{y}'_{i,t}(n)\mathbf{\Phi}_{i,t} + \mathbf{y}'_{i,t}(n)\left(\mathbf{\Phi}_{i,t} - \mathbf{\Psi}_{i,t}\right)\frac{\kappa_{i,t}}{1-\kappa_{i,t}}, \tag{21}$$

$$= \exp\left(\ln\left(U_{i,t}(n) \right) + \ln\left(1 + \left(1 - \frac{V_{i,t}(n)}{U_{i,t}(n)}\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \tag{22}$$

where κ is defined as

$$\kappa_{i,t} = \frac{\sum_{n=1}^{N} U_{i,t}(n)V_{i,t}(n)}{\gamma_{i,t} + \sum_{n=1}^{N} V_{i,t}^2(n)}. \tag{23}$$

Proof of Proposition 1. *See Appendix A.3.* □

Importantly, Proposition 1 shows that the characteristic-based asset demand can be flexibly derived from mean-variance portfolio choice without additional assumptions on the functional form for $U(\cdot)$ and $V(\cdot)$. Nonetheless, by imposing log-linearity of expected returns $U(\cdot)$ and factor loadings $V(\cdot)$ in asset characteristics $\mathbf{x}_{it}$, we can explore how specific modeling choices recover existing approaches in the demand system asset pricing literature, and discuss the relative merits of our model.

Assumption 5 (Log-Linear Specification of Expected Returns and Factor Loadings). *If* (19) *and* (20) *are log-linear, we get*

$$U_{i,t}(n) = \exp\left(\mathbf{x}'_{i,t}\beta_{i,t} \right), \tag{24}$$

$$V_{i,t}(n) = \exp\left(\mathbf{x}'_{i,t}\eta_{i,t} \right). \tag{25}$$

This log-linear specification implicitly assumes that the unobserved component of asset characteristics affects both expected returns $U(\cdot)$ and factor loadings $V(\cdot)$ proportionally. However, this assumption can be relaxed to allow for heterogeneity in how $U(\cdot)$ and

$V(\cdot)$ depend on unobserved demand drivers. For example, introducing different powers of the unobserved component, e.g., $\varepsilon_{i,t}(n)$ in $U(\cdot)$ and $[\varepsilon_{i,t}(n)]^{\tau}$ with $\tau \neq 1$ in $V(\cdot)$ enables the model to capture varying sensitivities to latent demand factors across systematic risk and return components. Furthermore, we can normalize the Equations (24) and (25) in the following lemma, allowing the intercept term to represent the demand for all assets relative to the outside asset in the investment universe.

Lemma 3 (Intercept Term in the Asset Demand System). *The intercept term indexed by K in the asset demand system satisfies*

$$\beta_{K,i,t} = \eta_{K,i,t} = -\ln\left(\,\varepsilon_{i,t}(0)\,\right). \tag{26}$$

Proof of Lemma 3. *See Appendix A.4.* □

With Lemma 3, our demand system aligns with economic intuition about investor preferences, highlighting the role of the outside asset as a benchmark for allocation, with the intercept capturing systematic influences on all assets. Building on the flexibility introduced by Proposition 1 and leveraging Assumption 5 and Lemma 3, the following corollary imposes structure on our generalized MNL specification with potential non-linearities. Despite these functional form assumptions, this parametrization retains the flexibility to capture rich substitution patterns, potential complementarities, and varying elasticities of expected returns and factor loadings with respect to asset characteristics.

Corollary 1 (Implications of Log-Linearity for the Asset Demand System). *When both $U(\cdot)$ and $V(\cdot)$ are log-linear and multiplicative in observed and unobserved components* *in $x_{i,t}(n)$ as in* (24) *and* (25)*, the generalized characteristic-based asset demand* (22) *becomes:*

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp\left(\mathbf{x}'_{i,t}(n)\beta_{i,t}\right)\left[1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right], \tag{27}$$

where we defined the difference in coefficients as $\Delta_{i,t} = (\eta_{0,i,t} - \beta_{0,i,t}, \dots, \eta_{K-1,i,t} - \beta_{K-1,i,t}, 0)$.

Proof of Corollary 1. *See Appendix A.5.* □

Thus, the term $\Delta_{i,t}$ permits variation in the coefficients for expected returns and factor loadings. However, to better align with existing approaches in the demand-systems asset pricing literature, additional symmetry assumptions should introduced. The following corollary formalizes this relationship and identifies the necessary and sufficient conditions under which the nonlinear MNL framework collapses into a linear MNL.

Corollary 2 (Necessary and Sufficient Condition for Linear MNL Specification). *Along with the log-linearity of $U(\cdot)$, the restriction*

$$\mathbf{\Phi}_{i,t} = \mathbf{\Psi}_{i,t}, \tag{28}$$

is a sufficient and necessary condition for deriving a linear MNL

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp\left(\mathbf{x}'_{i,t}(n)\beta_{i,t}\right), \tag{29}$$

from the non-linear MNL (22).

Proof of Corollary 2. *See Appendix A.6.* □

Under Corollary 2, the log-linear demand system simplifies to the linear MNL specification proposed by KY19. This restriction, while computationally convenient — eliminating all nonlinear terms — imposes proportional responses of expected returns and factor loadings to asset characteristics. Notably, it implies that the single factor considered by the investor fully prices all assets, leaving no room for alpha.

We close our model with market clearing for each asset n:

$$\mathbf{ME}_t(n) = \sum_{i=1}^{I} A_{i,t} w_{i,t}(n),$$

where the market value of shares outstanding must equal the wealth-weighted sum of portfolio weights across all investors.

2.4. Implications of Cross-Asset Dependencies

To provide further context, we revisit the example introduced earlier, interpreting this result within a specific scenario relative to our generalized demand system.

Example. *Assume that Alice and Bob believe institutions consider only value and size strategies. As a result, they form expectations about returns and factor loadings based solely on two characteristics: log book-to-market equity (value) and log market equity (size). Any changes in their risk compensation during market events can be modeled using the coefficients on asset characteristics in Equation 21:*

$$\mathbf{\Phi}_{i,t} + (\mathbf{\Phi}_{i,t} - \mathbf{\Psi}_{i,t}) \frac{\kappa_{i,t}}{1 - \kappa_{i,t}}$$

Alice holds shares of Pfizer, a value stock. Recently, Pfizer announced increased retained earnings for future drug development, boosting its book equity without a corresponding increase in market equity. As the book-to-market ratio rises, Alice expects the stock to become more attractive to value-oriented investors. This change increases Pfizer's expected returns' sensitivity to *log book equity with its sensitivity to log market equity unchanged. To keep the same weights:*

$$\Phi_{Alice,t}(be_t(Pfizer)) \quad \to \quad (1+10\%) \times \Phi_{Alice,t}(be_t(Pfizer))$$

$$\Psi_{Alice,t}(be_t(Pfizer)) \quad \to \quad \left(1 + \frac{\Phi_{Alice,t}(be_t(Pfizer))}{\Psi_{Alice,t}(be_t(Pfizer))} \times \frac{1}{\kappa_{Alice,t}} \times 10\%\right) \times \Psi_{Alice,t}(be_t(Pfizer))$$

Bob holds shares of Etsy, a small-cap stock. Last year, Etsy introduced a premium subscription service offering tools to help sellers grow their businesses, such as sponsored product credits, enhanced ad placements, and detailed buyer insights. This quarter, Etsy reported higher revenues from subscription fees, increasing its market equity. Bob anticipates that Etsy's expected returns' sensitivity to log market equity will rise. To maintain the same portfolio weights:

$$\Phi_{Bob,t}(me_t(Etsy)) \quad \to \quad (1+10\%) \times \Phi_{Bob,t}(me_t(Etsy))$$

$$\Psi_{Bob,t}(me_t(Etsy)) \quad \to \quad \left(1 + \frac{\Phi_{Bob,t}(me_t(Etsy))}{\Psi_{Bob,t}(me_t(Etsy))} \times \frac{1}{\kappa_{Bob,t}} \times 10\%\right) \times \Psi_{Bob,t}(me_t(Etsy))$$

However, if Alice's and Bob's asset demand were modeled using a linear MNL, the restriction $\boldsymbol{\Phi}_{i,t} = \boldsymbol{\Psi}_{i,t}$ *from Corollary 2 imposes that the model-implied required increase in risk for a given increase in expected returns would be uniform across characteristics:*

$$\Psi_{Alice,t}(be_t(Pfizer)) \quad \to \quad \left(1 + \frac{1}{\kappa_{Alice,t}} \times 10\%\right) \times \Psi_{Alice,t}(be_t(Pfizer))$$
$$\Psi_{Bob,t}(me_t(Etsy)) \quad \to \quad \left(1 + \frac{1}{\kappa_{Bob,t}} \times 10\%\right) \times \Psi_{Bob,t}(me_t(Etsy))$$

Since κ is a portfolio-specific parameter, the sensitivity of the factor loading on the market factor with respect to any asset characteristic would no longer depend on that specific characteristic.

This outcome aligns with the economic intuition of $\boldsymbol{\Phi}_{i,t} = \boldsymbol{\Psi}_{i,t}$, which implies that the market factor perfectly prices all assets. In such a scenario, characteristics lose their explanatory power for heterogeneity in risk-return profiles across institutional investors. Consequently, using characteristics to infer differences in risk-return trade-offs would be futile beyond predicting portfolio holdings.

This restriction is significant because there is likely heterogeneity in how different characteristics influence risk-return profiles. For example, the assets in Alice's portfolio may exhibit greater sensitivity to log book equity, potentially generating alpha relative to the CAPM. This suggests that the assumption of $\mathbf{\Phi}_{i,t} = \mathbf{\Psi}_{i,t}$ is overly strong and may fail to capture the nuanced heterogeneity across characteristics and investors.

Even with an imposed log-linear functional form, a key feature of our model is its ability to accommodate the substantial heterogeneity in risk-return profiles across institutional investors by allowing $\mathbf{\Phi}_{i,t} \neq \mathbf{\Psi}_{i,t}$. For instance, value investors may focus on book-to-market ratios, while momentum investors prioritize past returns, and green investors emphasize ESG metrics. The non-linear correction term in our model adjusts for the net risk-return trade-off in an investor's portfolio with respect to these differing characteristics, enabling it to capture the diversity of institutional preferences and portfolio dynamics across distinct investment strategies.

2.5. Price Elasticities of Asset Demand

Understanding how investors respond to price changes is central to asset pricing. Price elasticities quantify these responses, reflecting how investor demand adjusts when prices move. In our framework, elasticities depend on the entire portfolio composition, contrasting with the KY19 MNL model, where cross-elasticities are rigidly proportional and cannot accommodate varying asset roles or characteristic-driven heterogeneity. For example, an investor's Tesla holdings respond to a price change in Ford proportional to their Ford holdings, and their Ford holdings respond to a price change in Tesla proportional to their Tesla holdings.

By relaxing this proportional structure, our model recognizes that an investor's Tesla holdings might respond differently to a price change in Ford than vice versa, due to distinct asset characteristics and their interactions within the portfolio. A price change in Tesla may trigger larger adjustments than a similar change in Ford, reflecting their unique contributions to the investor's overall risk-return trade-off.

The source of this complexity lies in our non-linear correction term, which accounts for how characteristics influence both expected returns and factor loadings. The price elasticities in our framework depend on portfolio-specific factors such as the factor risk exposed returns $\kappa_{i,t}$, as well as individual asset characteristics through $U(\cdot)$ and $V(\cdot)$. The non-linear correction term in our model plays a central role, adjusting the elasticities to account for both the direct response to price changes and the portfolio-wide adjustments driven by cross-asset dependencies. This correction term ensures that elasticities are sensitive to the heterogeneity in how expected returns and factor loadings interact with the investor's portfolio composition, capturing the complexity of investment decisions.

The following corollary introduces the elasticity formula implied by our model and lays the foundation for interpreting its components, particularly the role of the correction term in shaping investor behavior:

Corollary 3 (Price Elasticity for an Individual Investor). *Let* $\mathbf{q}_{i,t} = \ln\left(A_{i,t}\mathbf{w}_{i,t} \right) - \mathbf{p}_t$ *be the vector of log shares held by investor* i*. Then, from Corollary 1, we derive the price* *elasticity of demand for an individual investor i as*

$$
\begin{aligned}
-\frac{\partial \mathbf{q}_{i,t}(n)}{\partial \mathbf{p}_t(j)} = & \underbrace{\mathbf{1}\{n=j\} - \beta_{0,i,t}\left[\mathbf{1}\{n=j\} - w_{i,t}(j)\right]}_{\text{Base Elasticities}} \\
& \underbrace{-\left[\mathbf{1}\{n=j\} + w_{i,t}(j)\right] \frac{\exp\left(\mathbf{x}'_{i,t}(j)\Delta_{i,t}\right)\kappa_{i,t}}{1-\exp\left(\mathbf{x}'_{i,t}(j)\Delta_{i,t}\right)\kappa_{i,t}}\Delta_{0,i,t}}_{\text{Adjustment for Net Risk Exposure}} \\
& \underbrace{-\frac{\partial \kappa_{i,t}}{\partial \mathbf{p}_t(j)} \sum_{r=0}^{N} w_{i,t}(r) \frac{\exp\left(\mathbf{x}'_{i,t}(r)\Delta_{i,t}\right) - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)}{(1-\kappa_{i,t}\exp\left(\mathbf{x}'_{i,t}(r)\Delta_{i,t}\right))(1-\kappa_{i,t}\exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right))}}_{\text{Adjustment for Cross-Asset Dependencies}}
\end{aligned}
\tag{30}
$$

The base elasticities measure how demand for an asset responds to price changes in a linear MNL framework as in KY19. In this elasticity component, only the price coefficient on expected returns matters, i.e., $\beta_{0,i,t}$, with no adjustments for a different sensitivity to factor loadings and the response is proportional to the portfolio weights. Thus, the role of the risk-return trade-off doesn't show up.

The second term shows how the differential sensitivities of expected returns and factor loadings to the characteristics, i.e., $\Delta_{i,t} \neq 0$, influence demand elasticities. This adjustment, further scaled by the market factor's contribution to the portfolio returns $\kappa_{i,t}$, effectively capturing how asset-specific characteristics impact the investor's portfolio rebalancing dynamics. For instance, if the net risk of market equity Δ_{0it} is negative, the second term becomes positive, making the price elasticity larger than that implied by KY19's MNL. Intuitively, this corresponds to a scenario where an asset's risk rises faster than its returns when its price increases, prompting the investor to cut back more aggressively. Consequently, this component ensures that elasticities accurately reflect the nuanced contributions of asset characteristics to both returns and systematic risk.

The third term introduces the complexity of cross-asset interactions by capturing how changes in one asset j's price affect the portfolio-wide exposure to systematic risk $\kappa_{i,t}$ and propagate across all other assets. It reflects the sensitivity of $\kappa_{i,t}$ to price changes and scales it by the relative differences in characteristics across assets. For example, assets with high portfolio weights or large coefficients on factor loadings exert a disproportionate influence on this term, amplifying their impact on other assets' elasticities. This component highlights the interconnected nature of portfolio optimization, where adjustments to one asset ripple through the entire portfolio, influencing the demand for seemingly unrelated assets. Similarly, if all assets share the same characteristics, the third term can also become positive under certain conditions, further amplifying the departure from the MNL benchmark and potentially inducing complementarity among assets.

3. Identification and Estimation

Having established the theoretical foundation of our demand system, we now turn to identifying and estimating its structural parameters. We detail how to bridge theory and empirical implementation by using a demand inverse approach, as in the empirical IO literature (Berry et al., 1995), to achieve parameter identification, and then applying a GMM procedure based on the generated moment conditions. We assume the researcher observes each asset's portfolio weights $(w_{i,t}(n))_{i=1}^{N_{i,t}}$ and characteristics $(x_t(n))_{n=1}^{N_{i,t}}$. While the discussion focuses on investor-date specific parameters, our approach extends naturally to settings where parameters are invariant across investors or time.

3.1. Identifying Structural Parameters

A core challenge in structural estimation of the characteristic-based demand (22) is understanding how observed portfolio weights relate to both individual asset characteristics and the broader, portfolio-wide risk-return environment. The first step centers on expressing the relative portfolio weights in a form that isolates the key structural parameters $(\beta_{it}, \sqrt{\gamma_{it}}, \Delta_{it})$. The following lemma provides the key initial step, rewriting the asset demand system in a way that consolidates the main parameters into a single functional form, and hence serving as the foundation for the subsequent identification results.

Lemma 4 (Reformulating Relative Weights for Demand Inversion). *The asset demand system* (22) *can be rewritten to isolate the key structural parameters:*

$$\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^2 = \underbrace{\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\frac{U_{i,t}(n)}{\sqrt{\gamma_{i,t}}}\right)}_{\tilde{U}_{i,t}(n)}\left[\sqrt{\gamma_{i,t}} + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\sum_{r=1}^{N}\tilde{U}_{i,t}(r)\exp\left(\mathbf{x}'_{i,t}(r)\Delta_{i,t}\right)\right],$$

$$:= F_n\left(\tilde{U}_{i,t}(n)\ ;\ \sqrt{\gamma_{i,t}}, \Delta_{i,t}\right), \qquad \forall n = 1, \ldots, N_{i,t}. \tag{31}$$

Proof of Lemma 4. *See Appendix A.7.* □

By expressing the portfolio weights in terms of $\tilde{U}_{i,t}(n)$, as in Lemma 4, we isolate the structural parameters into a neat functional relationship. This representation is crucial because it prepares the system for the next step: inverting the demand function to recover these parameters from observed data. To proceed, we must first ensure that this demand system can indeed be inverted. The next proposition establishes the local invertibility conditions, building directly on the functional form from Lemma 4 to show that we can map observable portfolio weights back to the underlying parameters $(\beta_{it}, \sqrt{\gamma_{it}}, \Delta_{it})$.

Proposition 2 (Invertibility of Asset Demand). *The asset demand can be inverted to*

$$\tilde{U}_{i,t}(n) = F_n^{-1}\left(\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^2 \; ; \; \sqrt{\gamma_{i,t}}, \Delta_{i,t}\right) = \frac{w_{i,t}(n)}{w_{i,t}(0)} \frac{U_{i,t}^0(n)}{\sqrt{\gamma_{i,t}^0}}, \qquad \forall n = 1, \ldots, N_{i,t}, \tag{32}$$

where $U_{i,t}^0(n)$ is $U_{i,t}(n)$ under the true values $(\beta_{i,t}, \sqrt{\gamma_{i,t}}, \Delta_{i,t}) = (\beta_{i,t}^0, \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0)$.

Proof of Proposition 2. *See Appendix A.8.* □

Proposition 2 shows that, given regularity conditions (such as a non-singular Jacobian of F_n), one can recover the structural parameters from the observed portfolio weights. This invertibility is not just a technical convenience; it is the cornerstone that allows us to connect theory with empirical implementation. Once we can invert the demand system, we can formulate moment conditions and proceed with estimation.

However, before estimating the model, it is critical to understand the precise source of identification for the parameters $(\beta_{i,t}, \sqrt{\gamma_{i,t}}, \Delta_{i,t})$. While Proposition 2 guarantees that inversion is possible, we still need to pinpoint what aspects of the data identify each parameter. The following corollary provides a first-order Taylor expansion around the true parameter values, clarifying how each parameter's variation leaves distinct traits in the observable data.

Corollary 4 (Source of Identification). *Given Lemma 4, the first-order Taylor expansion of* $F_n\left(\tilde{U}_{i,t}(n)\ ;\ 1, 0\right)$ *for* $n = 1, \ldots, N_{i,t}$ *around* $(\sqrt{\gamma_{i,t}}, \Delta_{i,t}) = (\sqrt{\gamma^0_{i,t}}, \Delta^0_{i,t})$ *is given by*

$$\ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) \approx \mathbf{me}_t(n)\beta^0_{0,i,t} + \mathbf{x}'_{-0,t}(n)\beta^0_{-0,i,t} + \ln\left(\frac{1}{\sqrt{\gamma^0_{i,t}}}\right) - U^0_{i,t}(n)\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^{-1}\left(\frac{1}{\sqrt{\gamma^0_{i,t}}} - 1\right)$$
$$+ \sum_{r=1}^{N_{i,t}} \frac{w_{i,t}(r)}{w_{i,t}(0)} V^0_{i,t}(r)\Big[\left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta^0_{i,t}\right)\right)\mathbf{x}'_{i,t}(r) - \exp\left(\mathbf{x}'_{i,t}(n)\Delta^0_{i,t}\right)\mathbf{x}'_{i,t}(n)\Big]\frac{\Delta^0_{i,t}}{\sqrt{\gamma^0_{i,t}}}$$
$$+ \ln\left(\epsilon_{i,t}(n)\right). \tag{33}$$

Proof of Corollary 4. *See Appendix A.9.* □

Corollary 4 shows how variation in portfolio weights and asset characteristics, combined with suitable, exogenous instruments, is sufficient to identify all structural parameters. Specifically, the price coefficient β^0_{0it} can still be recovered using plausibly exogenous investment mandates as instruments, as in Koijen and Yogo (2019). Exogenous variation in characteristics other than price, $\mathbf{x}_{-0,t}(n)$, identifies $\beta^0_{-0,i,t}$, while additional instruments are required to identify the parameters $\sqrt{\gamma^0_{i,t}}$ and $\Delta^0_{i,t}$. In essence, the corollary clarifies how each element of the model contributes to pin down the underlying parameters, thereby guiding the estimation strategy.

Investment mandates, as in Koijen and Yogo (2019), naturally serve as exogenous instruments for prices, given that they impose predetermined, persistent constraints on investors' investment universes, such as sector-specific funds or index-tracking mandates.[1] Although Fuchs et al. (2024) highlight that the exclusion restriction for such instruments might be violated under price spillovers, our model explicitly and separately incorporates

[1] KY19, Koijen et al. (2023) and Bretscher et al. (Forthcoming) show that institutional investors have persistent holdings and hence their investment universes are very stable over time.

these portfolio-wide interactions. By doing so, we show that the price coefficient $\beta^0_{0,i,t}$ can be identified by the argument of "cost shifters" despite complex cross-asset dependencies. To shed light on this point, consider the following instrument for $\text{me}_t(n)$ proposed by Koijen and Yogo (2019): for each asset $n = 1, ..., N$,

$$z^{\text{me}}_{it}(n) = \log\left(\sum_{j\neq i} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}\right).$$

Moreover, we have the market-clearing condition for each $n = 1$ at time t:

$$\text{me}_t(n) = \log\left(\sum_{i=1}^{I} A_{i,t}w_{i,t}(n)\right) - s_t(n), \tag{34}$$

where $s_t(n)$ is the log of the total number of shares of asset n at time t. First, under the assumption that both $A_{j,t}$ and $\mathcal{N}_j$ are predetermined and exogenous with respect to current demand shocks, $z^{\text{me}}_{it}(n)$ is therefore orthogonal to the demand shocks. Second, note that given i and t, the cross-asset variation in $z^{\text{me}}_{it}(n)$ is mainly driven by $\log\left(\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}\right)$,[2] the log of the frequency of the asset n in the investor's investment universe weighted by her wealth. Take two assets, say m and n, with similar non-price characteristics. If this frequency for asset m is higher than asset n, then m appears more often in the investor's universe than n, leading to greater overall investment in m than n (i.e., $\sum_{i=1}^{I} A_{i,t}w_{i,t}(m) > \sum_{i=1}^{I} A_{i,t}w_{i,t}(n)$) if $\text{me}_t(m) = \text{me}_t(n)$. Then, according to (34), one has to increase m's price relative to n's (i.e., a greater left-hand side of (34) for

[2]Note that $z^{\text{me}}_{it}(n) = \log\left(1 - \frac{\frac{A_{i,t}\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}}{\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}}\right) + \log\left(\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}\right)$. When $\frac{A_{i,t}\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}$ is small relative to $\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}$, the first term will be approximately $\frac{\frac{A_{i,t}\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}}{\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}}$ that is dominated by $\log\left(\sum_{j=1}^{N} A_{j,t}\frac{\mathbf{1}\{n\in\mathcal{N}_{j,t}\}}{1+|\mathcal{N}_{i,t}|}\right)$.

asset m) to clear the market for asset m. Intuitively, $z_{it}^{\mathrm{me}}(n)$ can be seen as a measure of asset n's popularity among investors. The more popular n is, the more the investor needs to pay to get one share, validating the desired first-stage predictive power of $z_{it}^{\mathrm{me}}(n)$ for $\mathrm{me}_t(n)$.

The assumption that non-price characteristics are exogenous is common in the demand system asset pricing literature that builds on KY19. Particularly, this assumption treats investors as atomistic, implying that their individual demand shocks do not shape aggregate conditions. However, as Kim (2024) highlights, correlated demand shocks driven by institutional rebalancing or procyclical risk-taking—can introduce factor structures into latent demand. Such correlations violate exogeneity and can bias elasticity estimates if not properly addressed.

Our framework also introduces the parameters $\sqrt{\gamma_{i,t}^0}$ and $\Delta_{i,t}^0$, which appear non-linearly in the regression via terms like $\frac{1}{\sqrt{\gamma_{it}^0}} - 1$ and $\frac{\Delta_{it}^0}{\sqrt{\gamma_{it}^0}}$. These parameters can be identified by a similar argument to the "BLP-type" instruments in the empirical IO literature (e.g., Berry et al. (1995); Gandhi and Houde (2019)). Concretely, we can use the exogenous characteristics of other assets in the investor's portfolio, $x_t(r)$, as valid instruments. These characteristics enter through

- the regressor for $\frac{1}{\sqrt{\gamma_{i,t}^0}} - 1$, provided that $\Delta_{it}^0 \neq 0$,
- the regressor for $\frac{\Delta_{it}^0}{\sqrt{\gamma_{i,t}^0}}$ via $x_t(r)$, $V_{i,t}^0(r)$, and $\frac{w_{i,t}(r)}{w_{i,t}(0)}$.

The resulting variation is non-linear and non-collinear, guaranteeing that we meet the rank conditions for separate identification of $(\gamma_{it}^0, \Delta_{it}^0)$. Moreover, under Lemma 3, i.e., normalizing the mean of latent demand to one, allows us to further identify the constants $\beta_{Kit}^0 + \log \frac{1}{\sqrt{\gamma_{it}^0}}$ and β_{Kit}^0, completing the identification of all structural parameters.

3.2. Generalized Method of Moments Estimation

Having established the identification strategy, we now turn to the estimation of structural parameters. We employ a GMM approach, leveraging the moment conditions implied by our model. To ensure these moment conditions are valid, we rely on the exogeneity of certain asset characteristics, as formalized in Assumption 6.

Assumption 6 (Exogeneity of Asset Characteristics). *There exists random variables $z_{i,t}(n)$ such that*

$$\mathbb{E}\left(\varepsilon_{i,t}(n) \mid z_{i,t}(n),\, x_{-0,t}(1), \ldots, x_{-0,t}(N_{i,t}) \right) = 1 \tag{35}$$

While KY19 assumes exogeneity only between an asset's own characteristics and its latent demand, Assumption 6 in addition accounts for cross-asset dependencies. This extension guarantees valid orthogonality conditions by requiring an asset's characteristics to be exogenous to other assets' latent demands. In doing so, it ensures error terms remain uncorrelated with the instruments, enabling consistent GMM estimation.

Building on this exogeneity assumption, Corollary 5 formalizes the moment conditions that link observed portfolio choices to the underlying structural parameters. These conditions isolate parameters through appropriate functions $g(\cdot)$. Consequently, they form the essential input for implementing a GMM estimator that recovers $(\beta^0_{it}, \sqrt{\gamma^0_{it}}, \Delta^0_{it})$.

Corollary 5 (Moment Conditions). *Given Proposition 2 and Assumption 6, we can construct unconditional moment condition for any measurable function $g\left(z_{i,t}(n), (x_{-0,t}(n))_{n=1}^{N_{i,t}}\right)$:*

$$\mathbb{E}\left(\left(\frac{F_n^{-1}\left(\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^2; \sqrt{\gamma_{i,t}}, \Delta_{i,t} \right)}{\exp\left(\mathbf{x}_t'(n)\beta^0_{i,t} \right)} \left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^{-1} \sqrt{\gamma^0_{i,t}} - 1 \right) g\left(z_{i,t}(n), x_{-0,t}(n)\right) \right) = 0, \tag{36}$$

for $n = 1, \ldots, N_{i,t}$.

Proof of Corollary 5. *See Appendix A.10.* □

Corollary 5 lays out the key orthogonality conditions that arise from combining the invertibility result with exogenous asset characteristics. These conditions express how transformations of the observed data must be uncorrelated with the error term, thereby enabling us to pin down the parameters. In essence, they provide the link from theory to an implementable estimation strategy.

Because we have $2K+2$ parameters, we need at least $2K+2$ such conditions to achieve identification. With sufficient cross-asset variation in instruments and characteristics, we can construct an appropriate set of $g(\cdot)$ functions that yield a non-singular system of equations. Using these conditions, we can apply GMM to estimate $(\beta^0_{it}, \sqrt{\gamma^0_{it}}, \Delta^0_{it})$.

3.3. Monte Carlo Simulations

Having established identification and laid out a GMM estimation approach, we now assess the performance of our estimator in a controlled setting. Monte Carlo simulations allow us to examine how well our method recovers the true structural parameters under known data-generating processes. By comparing our estimator's performance against both an oracle benchmark, which represents an upper bound under partial knowledge of the true structural parameters, and KY19's MNL approach, we can quantify the bias introduced by ignoring cross-asset dependencies.

In the simulations, we generate data under Proposition 1, i.e., from a characteristic-based demand that explicitly incorporates cross-asset dependencies, for a mean-variance investor holding a 500-asset portfolio. Next, we draw 200 random latent demands ε, and consider cost shifters to instrument for price endogeneity. Finally, we estimate the parameters using the three estimators: (i) oracle, (ii) KY19, and (iii) our estimator (AW24).

Table 2: Monte Carlo Simulations - Bias in Coefficients

This table shows that ignoring cross-asset dependencies introduces significant bias in asset demand estimations. A true model is generated under the assumption that a characteristic-based demand model with cross-asset dependencies is a good representation of how investors behave in financial markets. We consider a mean-variance investor with 500 assets in her portfolio, with the portfolio weights modeled taking into account cross-asset dependencies:

$$w_{i,t}(n) = \frac{\exp\left(\mathbf{x}'_{i,t}(n)\beta_{i,t}\right)\left[1+\left(1-\exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right]}{1+\sum_{r=1}^{N}\exp\left(\mathbf{x}'_{i,t}(r)\beta_{i,t}\right)\left[1+\left(1-\exp\left(\mathbf{x}'_{i,t}(r)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right]}$$

The investor considers the log market equity and another asset characteristic. The true model parameters are $\beta = (-1, 1, 0)$, $\Delta = (0.5, 0.5)$ and $\sqrt{\gamma} = \sqrt{2}$. Given this true model, we run 200 Monte Carlo simulations with different latent demand shocks ε. We consider cost shifters to instrument for price endogeneity: $me(n) = mc(n) + 2\mathbf{x}(n) + \varepsilon$. In each draw, we estimate the true parameters using (i) an oracle estimator where the $\kappa_{i,t}$ is known, (ii) the Koijen and Yogo (2019) estimator ignoring cross-asset dependencies, (iii) our estimator with cross-asset dependencies.

	Bias		
	Oracle	**KY19**	**AW24**
β_{me}	0.024	0.417	0.055
β_{other}	0.054	0.245	0.115
constant	0.036	0.299	0.101
$\sqrt{\gamma}$	0	-	0.073
Δ_{me}	0	-	0.090
Δ_{other}	0	-	0.126
	Median RMSE		
Own Price Elasticity	0.024	0.772	0.191

Looking at the results in Table 2, we see that ignoring cross-asset dependencies (KY19) introduces substantial bias in nearly all parameters, often an order of magnitude higher either of the other two estimators. Our estimator (AW24) reduces these biases markedly, though it does not completely eliminate them; nonetheless, it is very close to oracle benchmark which represents the upper bound. Notably, the median RMSE for own price elasticity is dramatically lower under our approach compared to KY19, indicating that accounting for cross-asset interactions greatly improves estimation accuracy.

4. Conclusion

This paper advances the literature on asset demand systems by developing a flexible characteristic-based framework that accommodates heterogeneous substitution patterns, potential cross-asset complementarities, and the pursuit of alpha through distinct investor strategies. By relaxing the rigid substitutability assumptions inherent in MNL models, our approach provides a more realistic and nuanced depiction of portfolio allocation behavior. This generalization bridges the gap between demand systems and mean-variance portfolio optimization, offering a robust framework capable of capturing diverse risk-return trade-offs and investor preferences across assets.

Beyond its theoretical advancements, we propose a demand inverse approach to identify structural parameters, linking the economic model to observed data via moment conditions, and then adopt a GMM procedure for estimation. To address price endogeneity, we validate the continued applicability of mandate-based instruments and extend the framework by incorporating the characteristics of other assets within an investor's portfolio as additional instruments to capture cross-asset dependencies. Monte Carlo simulations confirm the robustness of our framework, showing that it substantially reduces biases in elasticity estimates under conditions of cross-asset dependency, enhancing both precision and interpretability.

Our model is versatile and can be applied to a variety of asset classes, including equities, bonds, and currencies. Its flexibility allows researchers to capture the complex interplay between asset characteristics and investor preferences across different markets. By integrating flexible substitution, cross-asset complementarity, and alpha pursuit, this paper provides a robust foundation for advancing demand system approaches in asset pricing and improving our understanding of investor behavior and market outcomes.

- References -

Allen, J., J. Kastl, and M. Wittwer (2024): "Estimating Demand Systems with Bidding Data," *Working Paper*.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma (Forthcoming): "Institutional Corporate Bond Pricing," *Review of Financial Studies*.

Campbell, J. Y. and L. M. Viceira (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press.

Chaudhary, M., Z. Fu, and J. Li (2023): "Corporate Bond Multipliers: Substitutes Matter," *Working Paper*.

Crawford, G. S., R. S. Lee, M. D. Whinston, and A. Yurukoglu (2018): "The welfare effects of vertical integration in multichannel television markets," *Econometrica*, 86, 891–954.

Crawford, G. S. and A. Yurukoglu (2012): "The welfare effects of bundling in multichannel television markets," *American Economic Review*, 102, 643–685.

Darmouni, O., K. Siani, and K. Xiao (2024): "Nonbank Fragility in Credit Markets: Evidence from a Two-Layer Asset Demand System," *Working Paper*.

Davis, C., M. Kargar, and J. Li (2024): "Why is Stock-Level Demand Inelastic? A Portfolio Choice Approach," *Working Paper*.

Deaton, A. and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312–326.

Dubé, J.-P. (2004): "Multiple discreteness and product differentiation: Demand for carbonated soft drinks," *Marketing Science*, 23, 66–81.

Eren, E., A. Schrimpf, and D. Xia (2024): "The Demand for Government Debt," *Working Paper*.

Ershov, D., J.-W. Laliberté, M. Marcoux, and S. Orr (Forthcoming): "Estimating Complementarity with Large Choice Sets: An Application to Mergers," *RAND Journal of Economics*.

Fama, E. F. and K. R. French (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116, 1–22.

Fan, Y. (2013): "Ownership consolidation and product characteristics: A study of the US daily newspaper market," *American Economic Review*, 103, 1598–1628.

Fang, C. (2024): "Monetary Policy Amplification through Bond Fund Flows," *Working Paper*.

Fosgerau, M., J. Monardo, and A. de Palma (2024): "The Inverse Product Differentiation Logit Model," *American Economic Journal: Microeconomics*, 16, 329–70.

Fuchs, W. M., S. Fukuda, and D. Neuhann (2024): "Demand-System Asset Pricing: Theoretical Foundations," *Working Paper*.

Gabaix, X. and R. S. J. Koijen (2024): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *Working Paper*.

Gabaix, X., R. S. J. Koijen, F. Mainardi, S. Oh, and M. Yogo (2023): "Asset Demand of U.S. Households," *Working Paper*.

Gandhi, A. and J.-F. Houde (2019): "Measuring substitution patterns in differentiated-products industries," *NBER Working paper*.

Gentzkow, M. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97, 713–744.

Grzybowski, L. and F. Verboven (2016): "Substitution between fixed-line and mobile access: the role of complementarities," *Journal of Regulatory Economics*, 49, 113–151.

Haddad, V., P. Huebner, and E. Loualiche (Forthcoming): "How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing," *American Economic Review*.

Huebner, P. (2023): "The Making of Momentum: A Demand-System Perspective," *Working Paper*.

Iaria, A. and A. Wang (2020): "Identification and Estimation of Demand for Bundles," *Working Paper*.

——— (2024): "An Empirical Model of Quantity Discounts with Large Choice Sets," *Working Paper*.

Jansen, K. A., W. Li, and L. Schmid (2024): "Granular Treasury Demand with Arbitrageurs," *Working Paper*.

Jiang, Z., R. J. Richmond, and T. Zhang (2024): "A Portfolio Approach to Global Imbalances," *Journal of Finance*, 79, 2025–2076.

——— (Forthcoming): "Understanding the Strength of the Dollar," *Journal of Financial Economics*.

Kim, B. (2024): "Correlated Demand Shocks and Asset Pricing," *Working Paper*.

Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo (2021): "Inspecting the Mechanism of Quantitative Easing in the Euro Area," *Journal of Financial Economics*, 140, 1–20.

Koijen, R. S. J., R. J. Richmond, and M. Yogo (2023): "Which Investors Matter for Equity Valuations and Expected Returns?" *Review of Economic Studies*, 91, 2387–2424.

Koijen, R. S. J. and M. Yogo (2019): "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127, 1475–1515.

——— (2023): "Understanding the Ownership Structure of Corporate Bonds," *American Economic Review: Insights*, 5, 73–91.

——— (2024): "Exchange Rates and Asset Prices in a Global Demand System," *Working Paper*.

Liu, H., P. K. Chintagunta, and T. Zhu (2010): "Complementarities and the demand for home broadband internet services," *Marketing Science*, 29, 701–720.

Lopez-Lira, A. and N. L. Roussanov (2023): "Do Common Factors Really Explain the Cross-Section of Stock Returns?" *Working Paper*.

Mainardi, F. (2023): "A Demand-Based Approach to Short-Selling," *Working Paper*.

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7, 77–91.

Noh, D., S. Oh, and J. Song (2023): "Unpacking the Demand for Sustainable Equity Investing," *Working Paper*.

Siani, K. (2024): "Raising Bond Capital in Segmented Markets," *Working Paper*.

Thomassen, Ø., H. Smith, S. Seiler, and P. Schiraldi (2017): "Multi-category competition and market power: a model of supermarket pricing," *American Economic Review*, 107, 2308–2351.

van der Beck, P. and C. Jaunin (2023): "The Equity Market Implications of the Retail Investment Boom," *Working Paper*.

Wang, A. (2024): "Demand with Complementarity: A Market-Share Approach," *Working Paper*.

Appendix

A. Proofs

Proof of Lemma 1. The expected log utility over the investor's terminal wealth at time T can be written as

$$\mathbb{E}_{i,t}\left(\ln\left(A_{i,T}\right)\right) = \ln\left(A_{i,T}\right) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}\left(\ln\left(\frac{A_{i,s+1}}{A_{i,s}}\right)\right),$$
$$\approx \ln\left(A_{i,T}\right) + \mathbb{E}_{i,t}\left(r_{t+1}(0) + \mathbf{w}_{i,t}'\mu_{i,t} + \frac{\mathbf{w}_{i,t}'\Sigma_{i,t}\mathbf{w}_{i,t}}{2}\right), \quad \text{(A.1)}$$

with $\mu_{i,t}$ and $\Sigma_{i,t}$ denoting the excess log returns and the covariance matrix of log returns. Mind that the log return on the portfolio is not the same as the linear combination of log of individual assets. However, over a short time interval, we can use a second-order Taylor approximate, which results in an approximation of expected log utility around mean-variance utility. In the continuous-time limit, this approximation becomes an exact solution as discussed in detail in Campbell and Viceira (2002, Chapter 2).

Then, the Lagrangian and the first-order condition for the portfolio choice problem are

$$\mathcal{L}_{i,t} = \ln\left(A_{i,T}\right) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}\left(r_{t+1}(0) + \mathbf{w}_{i,t}'\mu_{i,t} + \frac{\mathbf{w}_{i,t}'\Sigma_{i,t}\mathbf{w}_{i,t}}{2} + \lambda_{i,s}\left(1 - \mathbf{1}'\mathbf{w}_{i,s}\right)\right),$$
$$\Rightarrow \quad \frac{\partial \mathcal{L}_{i,t}}{\partial \mathbf{w}_{i,t}} = \mu_{i,t} - \Sigma_{i,t}\mathbf{w}_{i,t} - \lambda_{i,t}\mathbf{1} = 0,$$
$$\Rightarrow \quad \mathbf{w}_{i,t} = \Sigma_{i,t}^{-1}\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1}\right). \quad \text{(A.2)}$$

Proof of Lemma 2. Under Assumption 1, the optimal portfolio weights from Lemma 1 are

$$\mathbf{w}_{i,t} = \Sigma_{i,t}^{-1}\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1}\right),$$

$$= \frac{1}{\gamma_{i,t}}\left(\mathbf{I} - \frac{\Gamma_{i,t}\Gamma'_{i,t}}{\Gamma'_{i,t}\Gamma_{i,t} + \gamma_{i,t}}\right)\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1}\right), \quad \Big| \quad \text{from the Woodbury matrix identity}$$

$$= \frac{1}{\gamma_{i,t}}\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1} - \Gamma_{i,t}\kappa_{i,t}\right), \quad \Big| \quad \kappa_{i,t} = \frac{\Gamma'_{i,t}\left(\mu_{i,t} - \lambda_{i,t}\mathbf{1}\right)}{\Gamma'_{i,t}\Gamma_{i,t} + \gamma_{i,t}}$$

$$= \pi_{i,t} + \mathbf{y}'_{i,t}(n)\mathbf{\Pi}_{i,t}, \quad \Big| \quad \begin{aligned} \pi_{i,t} &= (\phi_{i,t} - \lambda_{i,t} - \psi_{i,t}\kappa_{i,t})/\gamma_{i,t} \\ \mathbf{\Pi}_{i,t} &= (\mathbf{\Phi}_{i,t} - \mathbf{\Psi}_{i,t}\kappa_{i,t})/\gamma_{i,t} \end{aligned} \tag{A.3}$$

Proof of Proposition 1. Under Lemma 2 and Assumptions 2, 3 and 4, we can write

$$\frac{w_{it}(n)}{w_{it}(0)} = 1 + \mathbf{y}'_{i,t}(n)\frac{\mathbf{\Pi}_{it}}{\pi_{it}}, \quad \Big| \quad \text{Lemma 2 \& Assumption 2}$$

$$= 1 + \mathbf{y}'_{i,t}(n)\frac{\mathbf{\Phi}_{it} - \mathbf{\Psi}_{it}\kappa_{it}}{1 - \kappa_{it}}, \quad \Big| \quad \text{Assumptions 3 \& 4}$$

$$= 1 + \mathbf{y}'_{i,t}(n)\left(\mathbf{\Phi}_{it} + (\mathbf{\Phi}_{it} - \mathbf{\Psi}_{it})\left(\frac{\kappa_{it}}{1 - \kappa_{it}}\right)\right),$$

$$= 1 + \mathbf{y}'_{i,t}(n)\mathbf{\Phi}_{it} + \big(1 + \mathbf{y}'_{i,t}(n)\mathbf{\Phi}_{it} - \big(1 + \mathbf{y}'_{i,t}(n)\mathbf{\Psi}_{it}\big)\big)\left(\frac{\kappa_{it}}{1 - \kappa_{it}}\right),$$

$$= U_{i,t}(n) + (U_{i,t}(n) - V_{i,t}(n))\left(\frac{\kappa_{it}}{1 - \kappa_{it}}\right),$$

$$= U_{i,t}(n)\left(1 + \left(1 - \frac{V_{i,t}(n)}{U_{i,t}(n)}\right)\left(\frac{\kappa_{it}}{1 - \kappa_{it}}\right)\right) \tag{A.4}$$

Proof of Lemma 3. When $n = 0$:

$$U_{i,t}(0, ..., 0, 1, \ln(\varepsilon_{i,t}(0))) = \exp(\beta_{K,i,t} + \ln(\varepsilon_{i,t}(0))), \tag{A.5}$$

$$V_{i,t}(0, ..., 0, 1, \ln(\varepsilon_{i,t}(0))) = \exp(\eta_{K,i,t} + \ln(\varepsilon_{i,t}(0))), \tag{A.6}$$

then, we have for (21):

$$\begin{aligned}
\frac{w_{i,t}(0)}{w_{i,t}(0)} &= \exp\left(\ln(U_{i,t}(0)) + \ln\left(1 + \left(1 - \frac{V_{i,t}(0)}{U_{i,t}(0)}\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
\Leftrightarrow \qquad 1 &= \exp\left(\beta_{K,i,t} + \ln(\varepsilon_{i,t}(0)) + \ln\left(1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{K,i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
\Leftrightarrow \qquad 0 &= \beta_{K,i,t} + \ln(\varepsilon_{i,t}(0)) + \ln\left(1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{K,i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right), \\
\Leftrightarrow \qquad \beta_{K,i,t} &= -\ln(\varepsilon_{i,t}(0)), \qquad \text{(A.7)}
\end{aligned}$$

resulting from $\Delta_{K,i,t} = 0$ and which holds for any value of κ_{it}.

Proof of Corollary 1.

$$\begin{aligned}
\frac{w_{i,t}(n)}{w_{i,t}(0)} &= \exp\left(\ln(U_{i,t}(n)) + \ln\left(1 + \left(1 - \frac{V_{i,t}(n)}{U_{i,t}(n)}\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
&= \exp\left(\ln\left(\exp\left(\mathbf{x}'_{i,t}\beta_{i,t}\right)\right) + \ln\left(1 + \left(1 - \frac{\exp\left(\mathbf{x}'_{i,t}\eta_{i,t}\right)}{\exp\left(\mathbf{x}'_{i,t}\beta_{i,t}\right)}\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
&= \exp\left(\mathbf{x}'_{i,t}\beta_{i,t} + \ln\left(1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
&= \exp\left(\mathbf{x}'_{i,t}\beta_{i,t}\right)\exp\left(\ln\left(1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right)\right), \\
&= \exp\left(\mathbf{x}'_{i,t}(n)\beta_{i,t}\right)\left[1 + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t}\right)\right)\left(\frac{\kappa_{i,t}}{1-\kappa_{i,t}}\right)\right]. \qquad \text{(A.8)}
\end{aligned}$$

Proof of Corollary 2. One needs two conditions to derive (29) in which β_{it} does not depend on asset characteristics $\mathbf{x}_t = (x_t(1), \ldots, x_t(N))$.

Condition 1. $\dfrac{\mathbf{\Pi}_{it}}{\pi_{it}}$ is a constant vector that does not depend on $\boldsymbol{x}_{it}$.

Condition 2. the coefficients in the constant vector satisfy the restrictions that these coefficients coincide with those of the Taylor expansion of the exponential function around zero.

Condition 2 is a functional form assumption on the asset-specific index and it is satisfied under Assumption 5. We show that **Condition 1** is the key to delivering the multinomial logit specification (and therefore the IIA restriction); note that from the proof of Proposition 1:

$$\frac{\mathbf{\Pi}_{it}}{\pi_{it}} = \mathbf{\Phi}_{it} + (\mathbf{\Phi}_{it} - \mathbf{\Psi}_{it}) \left(\frac{\kappa_{it}}{1-\kappa_{it}} \right), \tag{A.9}$$

which is constant if and only if when $\kappa_{it} = 0$ or $\mathbf{\Phi}_{it} = \mathbf{\Psi}_{it}$. However, note that $\kappa_{i,t} > 0$ because $U(\cdot)$ and $V(\cdot)$ are both positive functions under Assumption 5. Thus, $\kappa_{i,t}$ can never be zero and $\dfrac{\mathbf{\Pi}_{it}}{\pi_{it}}$ is a constant vector if and only if $\mathbf{\Phi}_{it} = \mathbf{\Psi}_{it}$.

Proof of Lemma 4. Given

$$\begin{aligned}
\sum_{n=1}^{N_{i,t}} \frac{w_{i,t}(n)}{w_{i,t}(0)} V_{i,t}(n) &= \sum_{n=1}^{N_{i,t}} U_{i,t}(n)V_{i,t}(n) + \sum_{n=1}^{N_{i,t}} V_{i,t}(n)\left(U_{i,t}(n) - V_{i,t}(n)\right) \frac{\kappa_{i,t}}{1-\kappa_{i,t}}, \\
&= \sum_{n=1}^{N_{i,t}} U_{i,t}(n)V_{i,t}(n) + \left(\sum_{n=1}^{N_{i,t}} U_{i,t}(n)V_{i,t}(n) - \sum_{n=1}^{N_{i,t}} V_{i,t}^2(n) \right) \frac{\sum_{n=1}^{N_{i,t}} U_{i,t}(n)V_{i,t}(n)}{\gamma_{i,t} + \sum_{n=1}^{N_{i,t}} V_{i,t}^2(n) - \sum_{n=1}^{N_{i,t}} U_{i,t}(n)V_{i,t}(n)}, \\
&= \gamma_{i,t} \frac{\kappa_{i,t}}{1-\kappa_{i,t}},
\end{aligned} \tag{A.10}$$

we can rewrite (22) as:

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = U_{i,t}(n) \left[1 + \frac{1}{\gamma_{i,t}} \left(1 - \frac{V_{i,t}(n)}{U_{i,t}(n)} \right) \sum_{r=1}^{N} \frac{w_{i,t}(r)}{w_{i,t}(0)} V_{i,t}(r) \right], \tag{A.11}$$

or equivalently,

$$\begin{aligned}
\left(\frac{w_{i,t}(n)}{w_{i,t}(0)} \right)^2 &= \underbrace{\left(\frac{w_{i,t}(n)}{w_{i,t}(0)} \frac{U_{i,t}(n)}{\sqrt{\gamma_{i,t}}} \right)}_{\tilde{U}_{i,t}(n)} \left[\sqrt{\gamma_{i,t}} + \left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta_{i,t} \right) \right) \sum_{r=1}^{N} \tilde{U}_{i,t}(r) \exp\left(\mathbf{x}'_{i,t}(r)\Delta_{i,t} \right) \right], \\
&:= F_n \left(\left(\tilde{U}_{i,t}(n) \right)_{n=1}^{N_{i,t}} \; ; \; \sqrt{\gamma_{i,t}}, \Delta_{i,t} \right),
\end{aligned} \tag{A.12}$$

where the right-hand side defines a function of $(\tilde{U}_{it}(n))_{n=1}^{N_{it}}$ given $(\sqrt{\gamma_{it}}, \Delta_{it})$.

Proof of Proposition 2. By the inverse function theorem

$$\left(F_n\left((\tilde{U}_{it}(n))_{n=1}^{N_{it}}; \sqrt{\gamma_{it}^0}, \Delta_{it}^0\right)\right)_{n=1}^{N_{it}} \tag{A.13}$$

is at least locally invertible at

$$(\tilde{U}_{it}(n))_{n=1}^{N_{it}} = \left(\frac{w_{it}(n)}{w_{it}(0)}\frac{U_{it0}(n)}{\sqrt{\gamma_{it}^0}}\right)_{n=1}^{N_{it}} \tag{A.14}$$

if the Jacobian matrix has non-zero determinant. Note that the Jacobian matrix is

$$\partial_{\tilde{U}}\left(F_n\left(\left(\tilde{U}_{i,t}(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{it}^0\right)\right)_{n=1}^{N_{i,t}}$$

$$= \sqrt{\gamma_{i,t}^0}\text{diag}\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)_{n=1}^{N_{i,t}}\left[\mathbf{I} + \frac{1}{\gamma_{i,t}^0}\begin{bmatrix} U_{i,t}^0(1) - V_{i,t}^0(1) \\ \vdots \\ U_{i,t}^0(N) - V_{i,t}^0(N)\end{bmatrix}\begin{bmatrix} V_{i,t}^0(1) \cdots V_{i,t}^0(N)\end{bmatrix}\right] \text{diag}\left(\left(U_{it0}^0(n)\right)^{-1}\right)_{n=1}^{N_{it}}, \tag{A.15}$$

Using the matrix determinant lemma, its determinant is non-zero if and only if

$$\det\left(\mathbf{I} + \frac{1}{\gamma_{i,t}^0}\begin{bmatrix} U_{i,t}^0(1) - V_{i,t}^0(1) \\ \vdots \\ U_{i,t}^0(N) - V_{i,t}^0(N)\end{bmatrix}\begin{bmatrix} V_{i,t}^0(1) \cdots V_{i,t}^0(N)\end{bmatrix}\right) = 1 + \frac{\sum_{n=1}^{N_{i,t}} V_{i,t}^0(n)\left(U_{i,t}^0(n) - V_{i,t}^0(n)\right)}{\gamma_{i,t}^0} \neq 0. \tag{A.16}$$

When the joint distribution of the latent demands $(\epsilon_{it}(1), ..., \epsilon_{it}(N))$ is continuous,

$$\sum_{n=1}^{N_{i,t}} V_{i,t}^0(n)\left(U_{i,t}^0(n) - V_{i,t}^0(n)\right) \tag{A.17}$$

is a continuous random variable conditional on $(x_{0,t}(n), x_{-0,t}(n))_{n=1}^{N_{i,t}}$ with $x_{-0,t}(n)$ the log market equity and $x_{-0,t}(n)$ any other asset characteristics, while $\gamma_{i,t}^0$ is a constant. Consequently,

$$\gamma_{i,t}^0 \neq \sum_{n=1}^{N_{i,t}} V_{i,t}^0(n) \left(U_{i,t}^0(n) - V_{i,t}^0(n)\right) \tag{A.18}$$

holds with probability one and thus

$$\det\left(\partial_{\tilde{U}}\left(F_n\left(\left(\tilde{U}_{i,t}(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{it}^0\right)\right)_{n=1}^{N_{i,t}}\right) \neq 0, \tag{A.19}$$

and the inverse holds locally.

Proof of Corollary 4. We fix $\tilde{U}_{i,t}(n)$ to $\tilde{U}_{i,t}^0(n)$ for $n = 1, ..., N_{i,t}$ and develop the first-order Taylor expansion of $F_n\left(\left(\tilde{U}_{i,t}(n)\right)_{n=1}^{N_{i,t}}\ ;\ 1, 0\right)$, around $(\sqrt{\gamma_{i,t}}, \Delta_{i,t}) = (\sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0)$:

$$\begin{aligned}
&\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) + \ln\left(\frac{U_{i,t}^0(n)}{\sqrt{\gamma_{i,t}^0}}\right) \\
&\quad = \log F_n\left(\left(\tilde{U}_{it0}(n)\right)_{n=1}^{N_{i,t}}\ ;\ 1, 0\right) \\
&\quad \approx \ln\left(F_n\left(\left(\tilde{U}_{i,t}^0(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0\right)\right) \qquad \Bigg| \qquad \text{Note that this is equal to } \ln\left(\left(\frac{w_{it}(n)}{w_{it}(0)}\right)^2\right) \\
&\qquad + \frac{\partial_{\sqrt{\gamma}} F_n\left(\left(\tilde{U}_{i,t}^0(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0\right)}{F_n\left(\left(\tilde{U}_{i,t}^0(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0\right)}\left(1 - \sqrt{\gamma_{i,t}^0}\right) + \frac{\partial_{\Delta_{i,t}^0} F_n\left(\left(\tilde{U}_{i,t}^0(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0\right)}{F_n\left(\left(\tilde{U}_{i,t}^0(n)\right)_{n=1}^{N_{i,t}}\ ;\ \sqrt{\gamma_{i,t}^0}, \Delta_{i,t}^0\right)}\left(-\Delta_{i,t}^0\right) \\
&\quad = 2\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) + U_{i,t}^0(n)\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^{-1}\left(\frac{1}{\sqrt{\gamma_{i,t}^0}} - 1\right) \\
&\qquad + \sum_{r=1}^{N_{i,t}} \frac{w_{i,t}(r)}{w_{i,t}(0)} V_{i,t}^0(r)\Big[\left(1 - \exp\left(\mathbf{x}_{i,t}'(n)\Delta_{i,t}^0\right)\right)\mathbf{x}_{i,t}'(r) - \exp\left(\mathbf{x}_{i,t}'(n)\Delta_{i,t}^0\right)\mathbf{x}_{i,t}'(n)\Big]\frac{\Delta_{i,t}^0}{\sqrt{\gamma_{i,t}^0}},
\end{aligned} \tag{A.20}$$

which is, for $n = 1, ..., N$, equivalent to

$$\ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) \approx \mathbf{me}_t(n)\beta^0_{0,i,t} + \mathbf{x}'_{-0,t}(n)\beta^0_{-0,i,t} + \ln\left(\frac{1}{\sqrt{\gamma^0_{i,t}}}\right) - U^0_{i,t}(n)\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^{-1}\left(\frac{1}{\sqrt{\gamma^0_{i,t}}} - 1\right)$$
$$+ \sum_{r=1}^{N_{i,t}} \frac{w_{i,t}(r)}{w_{i,t}(0)} V^0_{i,t}(r)\Big[\left(1 - \exp\left(\mathbf{x}'_{i,t}(n)\Delta^0_{i,t}\right)\right)\mathbf{x}'_{i,t}(r) - \exp\left(\mathbf{x}'_{i,t}(n)\Delta^0_{i,t}\right)\mathbf{x}'_{i,t}(n)\Big]\frac{\Delta^0_{i,t}}{\sqrt{\gamma^0_{i,t}}}$$
$$+ \ln\left(\epsilon_{i,t}(n)\right). \tag{A.21}$$

Proof of Corollary 5. Using Assumption 5 and Proposition 2, we get

$$\Rightarrow \quad \frac{w_{i,t}(n)}{w_{i,t}(0)}\frac{U^0_{i,t}(n)}{\sqrt{\gamma^0_{i,t}}} = F_n^{-1}\left(\left(\frac{w_{i,t}(1)}{w_{i,t}(0)}\right)^2, \ldots, \left(\frac{w_{i,t}(N_{i,t})}{w_{i,t}(0)}\right)^2 \;;\; \sqrt{\gamma_{i,t}}, \Delta_{i,t}\right) \quad \Bigg| \quad \text{Proposition 2}$$

$$\Leftrightarrow \quad U^0_{i,t}(n) = \frac{F_n^{-1}\left(\cdot\;;\;\sqrt{\gamma_{i,t}}, \Delta_{i,t}\right)\cdot w_{i,t}(0)\cdot\sqrt{\gamma^0_{i,t}}}{w_{i,t}(n)}$$

$$\Rightarrow \quad \exp\left(\mathbf{x}'_t(n)\beta^0_{i,t}\right)\cdot\varepsilon_{i,t} = \frac{F_n^{-1}\left(\cdot\;;\;\sqrt{\gamma_{i,t}}, \Delta_{i,t}\right)\cdot w_{i,t}(0)\cdot\sqrt{\gamma^0_{i,t}}}{w_{i,t}(n)} \quad \Bigg| \quad \text{Assumption 5}$$

$$\Leftrightarrow \quad \varepsilon_{i,t} = \frac{F_n^{-1}\left(\cdot\;;\;\sqrt{\gamma_{i,t}}, \Delta_{i,t}\right)\cdot w_{i,t}(0)\cdot\sqrt{\gamma^0_{i,t}}}{\exp\left(\mathbf{x}'_t(n)\beta^0_{i,t}\right)\cdot w_{i,t}(n)} \tag{A.22}$$

Substituting the expression for $\varepsilon_{i,t}$ in Assumption 6, we have

$$\mathbb{E}\left(\frac{F_n^{-1}\left(\cdot\;;\;\sqrt{\gamma_{i,t}}, \Delta_{i,t}\right)\cdot w_{i,t}(0)\cdot\sqrt{\gamma^0_{i,t}}}{\exp\left(\mathbf{x}'_t(n)\beta^0_{i,t}\right)\cdot w_{i,t}(n)}\;\middle|\; z_{i,t}(n),\, x_{-0,t}(1), \ldots, x_{-0,t}(N_{i,t})\right) = 1 \tag{A.23}$$

To create a valid moment condition, multiply both sides of the equation by any measurable function $g\left(z_{i,t}(n), x_{-0,t}(1), \ldots, x_{-0,t}(N_{i,t})\right)$ and take expectations over the entire distribution to get

$$\mathbb{E}\left(\left(\frac{F_n^{-1}\left(\left(\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right)^2\right)_{n=1}^{N_{i,t}}\;;\;\sqrt{\gamma_{i,t}}, \Delta_{i,t}\right)\cdot w_{i,t}(0)\cdot\sqrt{\gamma^0_{i,t}}}{\exp(\mathbf{x}'_t(n)\beta^0_{i,t})\cdot w_{i,t}(n)} - 1\right) g\left(z_{i,t}(n), \left(x_{-0,t}(n)\right)_{n=1}^{N_{i,t}}\right)\right) = 0 \tag{A.24}$$