Identification and estimation of dynamic random coefficient models

Wooyong Lee[*]

May 2, 2025

Abstract

I study panel data linear models with predetermined regressors (such as lagged dependent variables) where coefficients are individual-specific, allowing for heterogeneity in the effects of the regressors on the dependent variable. I show that the model is not point-identified in a short panel context but rather partially identified, and I characterize the identified sets for the mean, variance, and CDF of the coefficient distribution. This characterization is general, accommodating discrete, continuous, and unbounded data, and it leads to computationally tractable estimation and inference procedures. I apply the method to study lifecycle earnings dynamics among U.S. households using the Panel Study of Income Dynamics (PSID) dataset. The results suggest substantial unobserved heterogeneity in earnings persistence, implying that households face varying levels of earnings risk which, in turn, contribute to heterogeneity in their consumption and savings behaviors.

Keywords: panel data regression, lagged dependent variable, heterogeneous coefficients, partial identification.

[*] Economics Discipline Group, UTS Business School, University of Technology Sydney. Address: 14-28 Ultimo Road, Ultimo, NSW 2007, Australia. Tel: (02) 9514 3074. Email: wooyong.lee.econ@gmail.com

1 Introduction

Panel data linear models with predetermined regressors (e.g., lagged dependent variables) are widely used in empirical research (Arellano and Bond, 1991; Blundell and Bond, 1998). Many of these models incorporate fixed effects, which are individual-specific intercepts that account for unobserved heterogeneity in the levels of the dependent variable. Fixed effects provide a flexible means of controlling for such heterogeneity, facilitating empirical research such as evaluation of a public policy. Fixed effects models are well understood in the context of short panel data (i.e., panel data with a small number of periods).

In addition to heterogeneity in the levels of dependent variables, there is ample evidence that individuals exhibit unobserved heterogeneity in the effects of regressors on dependent variables. For example, firms have varying degrees of labor efficiency in production; individuals experience different returns on education; and households differ in the persistence of earnings with respect to their past earnings. Such heterogeneous effects are crucial mechanisms for generating heterogeneous responses to exogenous shocks and policies, such as employment subsidies, tuition assistance, and income tax reforms. Moreover, these heterogeneous effects play a first-order role in determining outcomes in various economic models. For instance, heterogeneity in earnings persistence drives differences in the earnings risk faced by households, which, in turn, influences their heterogeneous motives for precautionary savings within lifecycle consumption models.

This paper examines a panel data linear model with predetermined regressors that permits unobserved heterogeneity in both the effects of regressors and the levels (i.e., a dynamic random coefficient model) in a short panel context. Consider a stylized example:

$$Y_{it} = \beta_{i0} + \beta_{i1} Y_{i,t-1} + \varepsilon_{it},$$

where all variables are scalars and ε_{it} is uncorrelated with the current history of Y_{it} (up to time $t-1$) but may be correlated with its future values. In this model, both the coefficient β_{i1} and the intercept β_{i0} are individual-specific, capturing heterogeneity in the effects of regressors and the levels. Moreover, the inclusion of the lagged dependent variable $Y_{i,t-1}$ as a regressor makes it a dynamic model.

Analysis of this model is challenging in short panels, as it is impossible to learn about individual values of the β_i parameters with a small number of periods. An influential study by Chamberlain (1993), recently republished as Chamberlain (2022), showed that the mean of the β_is in dynamic random coefficient models is not point-identified, implying that it cannot be consistently estimated. Since this negative result in the 1990s, progress in the literature has been limited. Arellano and Bonhomme (2012) showed that, for binary regressors, the mean of the β_is for certain subpopulations is identifiable and thus consistently estimable, but they did not establish a general identification result applicable to non-binary regressors. Most research on random coefficient models in short panels has focused on non-dynamic contexts (Chamberlain, 1992; Wooldridge, 2005; Arellano and Bonhomme, 2012; Graham and Powell, 2012), which exclude important dynamic mechanisms, such as the feedback from the current dependent variable to future regressors. For instance, a firm's labor purchase decision in the following period may depend on its current output, as the firm might learn about its own labor efficiency from that output. Moreover, understanding these dynamic mechanisms is of interest in its own. For example, a household's earnings persistence with respect to its past earnings is an important parameter, since high persistence increases the duration of earnings shocks, diminishing the household's ability to smooth consumption and, ultimately, impacting welfare.

This paper is, to the best of my knowledge, the first to present a general identification result for dynamic random coefficient models in a short panel context. Identification results are presented for various features of the coefficients, including their mean, variance, and cumulative distribution function (CDF). In addition, this paper proposes a computationally feasible estimation and inference procedure for these features. The procedure is then applied to investigate unobserved heterogeneity in lifecycle earnings dynamics among U.S. households using the Panel Study of Income Dynamics (PSID) dataset. These are presented in three steps.

First, I show that dynamic random coefficient models are partially identified, and I characterize finite lower and upper bounds for a class of parameters including the mean, variance, and CDF of the coefficient distribution. While these characterizations yield bounds that are not necessarily sharp, they are sufficiently general to accommodate discrete, continuous, or unbounded data. Moreover, for the mean of the coefficient distribution, the characterization yields a simple closed-form expression for its bounds, which clearly demonstrates that the bounds remain finite even when the data are unbounded, provided that certain moments of the data are finite. These results are obtained by recasting the identification problem as a linear programming problem (Honoré and Tamer, 2006; Honoré and Lleras-Muney, 2006; Mogstad, Santos, and Torgovitsky, 2018; Torgovitsky, 2019), which becomes infinite-dimensional when the data or the coefficients are continuous. I then employ the dual representation of infinite-dimensional linear programming (Galichon and Henry, 2009; Schennach, 2014) to derive the bounds for the parameters of interest.

Second, I propose computationally efficient estimation and inference procedures for the bounds. For the mean of the coefficient distribution, the closed-form expressions for its lower and upper bounds yield a simple and easy-to-implement estimation and inference procedure. In particular, I adopt the approach of Stoye (2020) and develop a simple procedure for constructing confidence intervals that are not only valid but also robust to overidentification and model misspecification. For other features of the coefficient distribution, such as the variance and the CDF, I use the approach of Andrews and Shi (2017), which performs inference on a countable number of moment inequalities. Although this procedure is computationally more demanding than that for the mean parameters, it remains computationally feasible for inference on various features of the coefficients.

Third, I estimate a reduced-form lifecycle model of earnings dynamics. Lifecycle earnings processes are key inputs in various economic models, including those of lifecycle consumption dynamics (Hall and Mishkin, 1982; Blundell, Pistaferri, and Preston, 2008; Blundell, Pistaferri, and Saporta-Eksten, 2016; Arellano, Blundell, and Bonhomme, 2017). Specifying an earnings process that captures features of real data is important for calibrating and drawing conclusions from these models. I investigate unobserved heterogeneity in the earnings of U.S. households using the Panel Study of Income Dynamics (PSID) dataset. Guvenen (2007, 2009) pointed out that, when allowing for unobserved heterogeneity in the time trend of earnings (known as a heterogeneous income profile, HIP), the estimated persistence of the income process is significantly below 1, with the latter being the estimate from the model that assumes no heterogeneity in the time trend (known as a restricted income profile, RIP). I extend this analysis by estimating a more general model that also permits unobserved heterogeneity in earnings persistence itself. I find that both the HIP and RIP specifications yield similar estimates of the average earnings persistence, with values significantly below 1. This suggests that misspecifying HIP as RIP (or vice versa) may not lead to serious model misspecification when earnings persistence is allowed to vary across households. Moreover, I find evidence of substantial unobserved heterogeneity in earnings persistence itself, implying that households face different levels of earnings risk, which in turn contributes to heterogeneity in their consumption and savings behavior.

The identification results in this paper can be extended to other structural models to accommodate heterogeneous effects. For example, these results can be applied to models with individual-specific coefficients and intercepts in probit and logit regressions. They can also be extended to vector-valued regressions, such as panel data vector autoregressive (VAR) models and systems of panel data regressions.

The remainder of this paper is structured as follows. Section 2 introduces the dynamic random coefficient model. Sections 3 and 4 present the identification results for the model, focusing on the mean in Section 3 and on more general features in Section 4. Section 5 discusses the estimation and inference procedures, and Section 6 applies these methods to lifecycle earnings dynamics. Section 7 concludes the paper. All proofs are provided in Online Appendix A.

2 Model and motivating examples

The dynamic random coefficient model is specified as follows:

$$Y_{it} = Z_{it}'\gamma_i + X_{it}'\beta_i + \varepsilon_{it}, \qquad t = 1, \ldots, T,$$

where i is an index of individuals, T is the length of panel data, $(Y_{it}, Z_{it}, X_{it}) \in \mathbb{R} \times \mathbb{R}^q \times \mathbb{R}^p$ are observed real vectors at time $t \in \{1, \ldots, T\}$, and $\varepsilon_{it} \in \mathbb{R}$ is the idiosyncratic error at time t. Let $R_{it} = (Z_{it}', X_{it}')'$ be the vector of regressors at time t, and let $B_i = (\gamma_i', \beta_i')'$ be the vector of random coefficients. With these definitions, I concisely write the model as:

$$Y_{it} = R_{it}'B_i + \varepsilon_{it}, \quad t = 1, \ldots, T. \tag{1}$$

Let $Y_i = (Y_{i1}, \ldots, Y_{iT})$ be the full history of $\{Y_{it}\}$, and let $Y_i^t = (Y_{i1}, \ldots, Y_{it})$ be the history of $\{Y_{it}\}$ up to time t. Define X_i, X_i^t, Z_i, Z_i^t similarly. I assume that

$$\mathbb{E}(\varepsilon_{it} | B_i, Z_i, X_i^t) = 0, \tag{2}$$

which implies that the error term is mean-independent of the full history of $\{Z_{it}\}$ (i.e., strict exogeneity) and of the current history of $\{X_{it}\}$ (i.e., sequential exogeneity). The inclusion of a sequentially exogenous regressor $\{X_{it}\}$ makes (1) a dynamic model. For example, the lagged dependent variable $Y_{i,t-1}$ can be included in X_{it}.

The model is studied in a short panel context, which corresponds to the asymptotics that the number of individuals $N \to \infty$ while the number of time periods T remains fixed. The random coefficients γ_i and β_i are unobserved random variables that follow nonparametric distributions, and they may be arbitrarily correlated with each other as well as with (Z_i, X_{i1}). This is how the random coefficient model extends a fixed effects model.

I summarize the variables of the model as two random vectors: the observable data $W_i = (Y_i', Z_i', X_i')' \in \mathcal{W}$ and the unobservable random coefficients $B_i \in \mathcal{B}$. I interpret ε_{it} as a deterministic function of (W_i, B_i) by the relationship $\varepsilon_{it} = Y_{it} - R_{it}'B_i$.

Given this model, I consider a parameter θ of the form

$$\theta = \mathbb{E}(m(Y_i, Z_i, X_i, \gamma_i, \beta_i)) = \mathbb{E}(m(W_i, B_i))$$

for some known function m. I present identification results for a generic function m, but I focus on the case in which m is either a polynomial or an indicator function of B_i, which allows for computationally feasible estimation and inference. This choice of m includes many important parameters of interest. For example, θ can be an element of the mean of the random coefficients $\mathbb{E}(B_i)$ or an element of the second moments $\mathbb{E}(B_i B_i')$. It can also represent the error variance $\mathbb{E}(\varepsilon_{it}^2)$ because $\varepsilon_{it}^2 = (Y_{it} - R_{it}' B_i)^2$ is a quadratic polynomial in B_i. Another example is the CDF of B_i evaluated at b, in which case one sets $m = \mathbf{1}(B_i \leq b)$ so that $\theta = \mathbb{E}(\mathbf{1}(B_i \leq b)) = \mathbb{P}(B_i \leq b)$.

Example 1 (Household earnings)**.** One of the simplest examples of (1) is the AR(1) model with heterogeneous coefficients:

$$Y_{it} = \gamma_i + \beta_i Y_{i,t-1} + \varepsilon_{it}, \tag{3}$$

where all variables are scalars. This is a special case of (1), with $Z_{it} = 1$ and $X_{it} = Y_{i,t-1}$.

The AR(1) model is a popular choice for empirical specification of the lifecycle earnings process, with Y_{it} representing log-earnings, an important input in models of consumption and savings behavior[1]. The earnings persistence parameter, β_i, governs the earnings risk experienced by households, which is a fundamental motive for precautionary savings. Specifying an earnings process that highlights features of real data is important for drawing conclusions from models of consumption and savings behavior. In the literature, the earnings process is often modeled as an AR(1) process with homogeneous coefficients (Lillard and Weiss, 1979; Blundell, Low, and Preston, 2013; Gu and Koenker, 2017), or as a unit root process, i.e., an AR(1) model with $\gamma_i = 0$ and $\beta_i = 1$ (Hall and Mishkin, 1982; Meghir and Pistaferri, 2004; Kaplan and Violante, 2014).

Example 2 (Household consumption behavior)**.** Consider a model of lifecycle consumption behavior:

$$C_{it} = \gamma_{i0} + \gamma_{i1} Y_{it} + \beta_i A_{it} + \nu_{it}, \tag{4}$$

where all variables are scalars. In this equation, C_{it} is non-durable consumption, Y_{it} is earnings, and A_{it} is asset holdings at time t, all measured in logs. In this specification, Y_{it} can be regarded as strictly exogenous, implying that future earnings are unaffected by the

[1] In the literature, it is standard to add a transitory shock to (3).

current consumption choice. In contrast, A_{it} must be taken as sequentially exogenous, as assets and consumption are interrelated through the intertemporal budget constraint.

The model in (4) can be viewed as an approximation of the consumption rule derived from a structural model (Blundell, Pistaferri, and Saporta-Eksten, 2016). One parameter of interest is γ_{i1}, the elasticity of consumption with respect to earnings. This elasticity measures a household's ability to smooth consumption in response to exogenous changes in earnings, such as earnings shocks, thereby mitigating adverse impacts on household welfare. Another parameter of interest is β_i, the elasticity of consumption with respect to asset holdings, which measures the household's capacity to smooth consumption in response to exogenous asset changes. Note that the model in (4) remains agnostic about the evolution of assets over time, i.e., it allows for a nonparametric evolution of the asset process.

Example 3 (Production function)**.** An influential paper by Olley and Pakes (1996) considered the estimation of the production function for firms operating with Cobb-Douglas technology. In their work, the following model was analyzed:

$$Y_{it} = \gamma_0 + \gamma_a A_{it} + \gamma_k K_{it} + \gamma_l L_{it} + \omega_{it} + \varepsilon_{it}$$

where Y_{it} is the log-output of firm i at time t, A_{it} is the firm's age, and K_{it} and L_{it} are the logs of capital and labor inputs, respectively. In this model, ω_{it} and ε_{it} are productivity shocks that are unobservable to the econometrician, while the firm observes ω_{it}.

Olley and Pakes (1996) assume that firm i's investment, I_{it}, is a strictly increasing function of ω_{it}, so that $I_{it} = g_t(\omega_{it}, A_{it}, K_{it})$. They then invert this function to obtain $\omega_{it} = h_t(I_{it}, A_{it}, K_{it})$, where $h_t = g_t^{-1}(\cdot, A_{it}, K_{it})$. When h_t is specified as a series function of its arguments, for example, a linear function $h_t = h_I I_{it} + h_A A_{it} + h_K K_{it}$ (for simplicity, the coefficients do not vary with t), the production function becomes

$$Y_{it} = \gamma_0 + \tilde{\gamma}_a A_{it} + \tilde{\gamma}_k K_{it} + \gamma_L L_{it} + h_I I_{it} + \varepsilon_{it},$$

where $\tilde{\gamma}_a = \gamma_a + h_A$ and $\tilde{\gamma}_k = \gamma_k + h_K$. Olley and Pakes (1996) then exploit additional moment restrictions implied by the model to separately identify (γ_a, γ_k) and (h_A, h_K). This approach has been extended and generalized by Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015).

In a recent contribution, Kasahara, Schrimpf, and Suzuki (2023) estimated a version of this model using a finite mixture specification for $(\gamma_0, \gamma_a, \gamma_k, \gamma_L)$, where they found an evidence of heterogeneity in these coefficients.

This paper also considers an extension of (1) that also involves regressors with homogeneous coefficients. Let $M_{it} = (\mathsf{Z}_{it}', \mathsf{X}_{it}')' \in \mathbb{R}^{q_m + p_m}$ be another vector of regressors, where Z_{it} is strictly exogenous and X_{it} is sequentially exogenous. Consider the model

$$Y_{it} = R_{it}' B_i + M_{it}' \delta + \varepsilon_{it}, \qquad t = 1, \ldots, T, \tag{5}$$

where $\delta \in \mathbb{R}^{q_m + p_m}$ is an unknown parameter, and assume that

$$\mathbb{E}(\varepsilon_{it} | B_i, Z_i, X_i^t, \mathsf{Z}_i, \mathsf{X}_i^t) = 0. \tag{6}$$

While I consider (1) as the main model of interest, I will also discuss how the results extend to the model in (5) in the context of the mean parameters.

The results of this paper also extend to a multivariate version of (1), namely, a system of random coefficient models. For example, one can combine the models in (3) and (4) to develop a joint lifecycle model of earnings and consumption behavior. This multivariate model permits the coefficients from the two processes to freely correlate among themselves and with (Y_{i0}, A_{i1}), allowing for potential correlation between the earnings and consumption processes. A full description of the multivariate model is provided in Online Appendix B.1.

3 Identification of the mean parameters

This section and the next section present identification results for the dynamic random coefficient model defined in (1) and (2). This section focuses on the identification of the mean parameters, and the next section extends the results to a more general class of parameters. Consider the mean of the random coefficient distribution:

$$\mu_e = \mathbb{E}(e_\gamma' \gamma_i + e_\beta' \beta_i) = \mathbb{E}(e' B_i)$$

where e_γ and e_β are real-valued vectors chosen by the econometrician and $e = (e_\gamma', e_\beta')'$. For example, if $e_\gamma = 0$ and $e_\beta = (1, 0, \ldots, 0)'$, then μ_e is the mean of the first entry of β_i.

This section is organized into four subsections. In the first, I show that μ_e is generally not point-identified. In the second, I show that μ_e is non-trivially partially identified, by deriving closed-form expressions for its finite lower and upper bounds. The third subsection then extends these results to the model that also involves homogeneous coefficients. Lastly, the fourth subsection provides a numerical illustration on the size of the bounds presented in this section. The results presented in this section are special cases of the more general results discussed in the next section and in Online Appendix B.3.

3.1 Failure of point identification

This subsection shows that μ_e is generally not point-identified, by considering a specific example of (1) and showing that μ_e is not point-identified in that example.

The example considered is the AR(1) model with heterogeneous coefficients in which two waves are observed:

$$Y_{it} = \gamma_i + \beta_i Y_{i,t-1} + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it}|\gamma_i, \beta_i, Y_i^{t-1}) = 0, \qquad t = 1, 2. \tag{7}$$

The following proposition states that $\mathbb{E}(\beta_i)$ is not point-identified in this model, which implies that there exists no consistent estimator for $\mathbb{E}(\beta_i)$.

Proposition 1. *Consider the model defined in (7). Assume that $(Y_{i0}, Y_{i1}, Y_{i2}, \gamma_i, \beta_i) \in \mathcal{C}$, where $\mathcal{C}$ is a compact subset of $\mathbb{R}^5$. Also assume that $(Y_{i0}, Y_{i1}, Y_{i2}, \gamma_i, \beta_i)$ is absolutely continuous with respect to the Lebesgue measure and that its joint density is strictly positive on $\mathcal{C}$. Then, $\mathbb{E}(\beta_i)$ is not point-identified.*

Chamberlain (1993), recently republished as Chamberlain (2022), showed that $\mathbb{E}(\beta_i)$ is not point-identified in (7) when Y_{it} is discrete and ε_{it} is mean-independent of Y_i^{t-1}. Proposition 1 complements this result by showing that point identification also fails under stronger assumptions and with continuous data. The failure of point identification in both the discrete and continuous cases in (7) suggests that this is a general feature of dynamic random coefficient models.

An intuition for Proposition 1 is as follows. Taking the first difference of (7) gives

$$Y_{i2} - Y_{i1} = \beta_i(Y_{i1} - Y_{i0}) + \varepsilon_{i2} - \varepsilon_{i1}.$$

Since $\varepsilon_{i2} - \varepsilon_{i1}$ has zero mean conditional on $(\gamma_i, \beta_i, Y_{i0})$, I obtain

$$\mathbb{E}(Y_{i2} - Y_{i1}|\gamma_i, \beta_i, Y_{i0}) = \mathbb{E}(\beta_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}),$$

which can be rewritten as

$$\mathbb{E}(Y_{i2} - Y_{i1} - \beta_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}) = 0. \tag{8}$$

Now, consider a function $k(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$ that is orthogonal to $Y_{i1} - Y_{i0}$ conditional on

$(\gamma_i, \beta_i, Y_{i0})$, i.e.,

$$\mathbb{E}(k(\gamma_i, \beta_i, Y_{i0}, Y_{i1})(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}) = 0.$$

For example, in the proof of Proposition 1, I choose such a function k to be

$$k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) = 1 - \frac{\mathbb{E}(Y_{i1} - Y_{i0}|\gamma_i, \beta_i, Y_{i0})}{\mathbb{E}((Y_{i1} - Y_{i0})^2|\gamma_i, \beta_i, Y_{i0})}(Y_{i1} - Y_{i0}).$$

Then, it follows that (8) holds true even if the original random coefficients (γ_i, β_i) are replaced with the following modified random coefficients:

$$\begin{aligned}\tilde{\gamma}_i &= \gamma_i - Y_{i1}k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}), \\ \tilde{\beta}_i &= \beta_i + k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}).\end{aligned}$$

To see this, note first that

$$\begin{aligned}&\mathbb{E}(\tilde{\beta}_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}) \\ &= \mathbb{E}(\beta_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}) + \mathbb{E}(k(\gamma_i, \beta_i, Y_{i0}, Y_{i1})(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}) \\ &= \mathbb{E}(\beta_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0})\end{aligned}$$

by the orthogonality property of k. Note also that conditioning on $(\gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1})$ is equivalent to conditioning on $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$ since $(\tilde{\gamma}_i, \tilde{\beta}_i)$ is a deterministic function of $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$. Then, by the law of iterated expectations and by (8), it follows that (8) holds true for the modified random coefficients $(\tilde{\gamma}_i, \tilde{\beta}_i)$:

$$\begin{aligned}&\mathbb{E}(Y_{i2} - Y_{i1} - \tilde{\beta}_i(Y_{i1} - Y_{i0})|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\ &= \mathbb{E}(\mathbb{E}(Y_{i2} - Y_{i1} - \tilde{\beta}_i(Y_{i1} - Y_{i0})|\tilde{\gamma}_i, \tilde{\beta}_i, \gamma_i, \beta_i, Y_{i0}, Y_{i1})|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\ &= \mathbb{E}(\mathbb{E}(Y_{i2} - Y_{i1} - \tilde{\beta}_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0}, Y_{i1})|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\ &= \mathbb{E}(\mathbb{E}(Y_{i2} - Y_{i1} - \tilde{\beta}_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0})|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\ &= \mathbb{E}(\mathbb{E}(Y_{i2} - Y_{i1} - \beta_i(Y_{i1} - Y_{i0})|\gamma_i, \beta_i, Y_{i0})|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) = \mathbb{E}(0|\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) = 0.\end{aligned}$$

However, if the function k is chosen such that $\mathbb{E}(k(\gamma_i, \beta_i, Y_{i0}, Y_{i1})) \neq 0$, which is true for the choice of k above, it follows that $\mathbb{E}(\tilde{\beta}) \neq \mathbb{E}(\beta)$.

Another intuition for Proposition 1 follows from an alternative proof of Proposition 1, which uses that $\mathbb{E}(\beta_i)$ is point-identified if and only if there exists an unbiased estimator of β_i in the individual time series. I state this result as a separate lemma below, which follows as a corollary of the general result in Online Appendix B.3.

Lemma 1. *Suppose that the assumptions of Proposition 1 hold, and that the regularity conditions* *stated as Assumption 13 in Online Appendix B.3 hold. Then $\mathbb{E}(\beta_i)$ is point-identified if and only if there exists a function $S^*(Y_{i0}, Y_{i1}, Y_{i2})$, which is a linear functional on the space of finite and countably additive signed Borel measures that are absolutely continuous with respect to the Lebesgue measure, such that*

$$\mathbb{E}(S^*(Y_{i0}, Y_{i1}, Y_{i2})|\beta_i) = \beta_i$$

almost surely. When such S^* exists, $\mathbb{E}(\beta_i)$ is identified by $\mathbb{E}(\beta_i) = \mathbb{E}(S^*(Y_{i0}, Y_{i1}, Y_{i2}))$.

Proposition 1 can then be proved by showing that there is no unbiased estimator of β_i (see Online Appendix B.2). The intuition for Lemma 1 is as follows. Since the distribution of β_i is unrestricted, information about individual β_i can only be obtained from its own time series. In a long panel context, a time series estimator of β_i that is consistent as $T \to \infty$ would reliably provide such information. In a short panel context, however, such an estimator is not reliable because T is finite. Lemma 1 shows that a time series estimator that is unbiased for finite T is the only reliable source of information on β_i when it comes to point identification in short panels.

3.2 Partial identification

A natural question following the last subsection is whether the data are at all informative about $\mu_e = \mathbb{E}(e'B_i)$, or whether they provide no information. This subsection shows that the data are indeed informative about μ_e. I show that there exist finite bounds L and U such that

$$L \leq \mu_e \leq U$$

where L and U are estimable from the observed data.

To identify μ_e, I use unconditional moment restrictions that are implications of (2). It is known that the set of unconditional moment restrictions of the form

$$\mathbb{E}(g(B_i, Z_i, X_i^t)\varepsilon_{it}) = 0, \tag{9}$$

indexed by a suitable class of functions g, is equivalent to the conditional moment restriction in (2) (Bierens, 1990; Andrews and Shi, 2013). I choose the class of g to be the set of polynomial functions and select a finite subset of these functions. This yields a finite number of unconditional moment restrictions that are fixed in the asymptotics that $N \to \infty$. This finite set of unconditional moment restrictions contains less information than the full conditional moment restriction in (2), yielding an outer bound rather than the sharp bound, but it leads to estimation and inference procedures that are computationally tractable. In addition, the empirical application in Section 6 shows that this finite set is sufficiently restrictive to provide informative bounds. Partial identification results based on the full conditional moment restriction in (2) are presented in Online Appendix B.3.

I now study the identification of μ_e. Recall the dynamic random coefficient model defined in (1) and (2):

$$Y_{it} = R_{it}' B_i + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it} | B_i, Z_i, X_i^t) = 0, \qquad t = 1, \ldots, T,$$

where $R_{it} = (Z_{it}', X_{it}')'$. For brevity of notation, define

$$Y_i \equiv \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix} \quad \text{and} \quad R_i \equiv \begin{pmatrix} R_{i1}' \\ \vdots \\ R_{iT}' \end{pmatrix}$$

as a random vector and a random matrix stacking Y_{it} and R_{it}' rowwise across t, respectively. Consider the following assumptions:

Assumption 1. (Y_i, Z_i, X_i, B_i) satisfies (1) and (2).

Assumption 2. $R_i' R_i$ is positive definite with probability 1.

Assumption 1 states that the dynamic random coefficient model is correctly specified. Assumption 2 is a no-multicollinearity assumption imposed on the individual time series. This is stronger than the assumption that $\mathbb{E}(R_i' R_i)$ is positive definite, a common assumption in standard dynamic fixed effect models. A stronger assumption is required because B_i is individual-specific with an unrestricted distribution, and each B_i can only be learned from its own individual data[2].

I now state a theorem showing that μ_e is partially identified under Assumptions 1 and 2. This theorem is a special case of Theorem 2 presented in the next section. For brevity of notation, define

$$\widehat{B}_i = (R_i' R_i)^{-1} R_i' Y_i, \quad \text{and} \quad \widehat{B}_0 = \mathbb{E}(R_i' R_i)^{-1} \mathbb{E}(R_i' Y_i).$$

Theorem 1. *Suppose that Assumptions 1 and 2 hold. Then $L \leq \mu_e \leq U$ where*

$$[L, U] = \left[\mathcal{B}_R - \frac{1}{2}\sqrt{\mathcal{E}_R \mathcal{D}_R},\ \mathcal{B}_R + \frac{1}{2}\sqrt{\mathcal{E}_R \mathcal{D}_R} \right],$$

[2] Graham and Powell (2012) studied a violation of Assumption 2 in a non-dynamic context.

and

$$\begin{aligned}\mathcal{B}_R &= \frac{1}{2}e'\mathbb{E}(\widehat{B}_i) + \frac{1}{2}e'\widehat{B}_0,\\ \mathcal{E}_R &= e'\mathbb{E}((R_i'R_i)^{-1})e - e'\mathbb{E}(R_i'R_i)^{-1}e,\\ \mathcal{D}_R &= \mathbb{E}(Y_iR_i(R_i'R_i)^{-1}R_i'Y_i) - \mathbb{E}(Y_iR_i)\mathbb{E}(R_i'R_i)^{-1}\mathbb{E}(R_i'Y_i).\end{aligned}$$

In addition, $\mathcal{E}_R \geq 0$ and $\mathcal{D}_R \geq 0$, and each is equal to zero if and only if $(R_i'R_i)^{-1}e$ and $(R_i'R_i)^{-1}R_i'Y_i$ are degenerate across individuals, respectively.

Note that $\widehat{B}_i$ is the individual-specific OLS estimator of B_i from its individual time series, $\widehat{B}_0$ is the pooled OLS estimator obtained by considering B_i as constant across individuals, and $R_i'R_i$ is the squared design matrix of the individual time series.

The closed-form expressions in Theorem 1 provide intuition on when L and U are finite. In particular, L and U are finite even if (Y_i, R_i, B_i) are unbounded, as long as the moments involved in the expression for $[L, U]$ are finite — that is, $\mathbb{E}(R_i'R_i)$, $\mathbb{E}((R_i'R_i)^{-1})$, $\mathbb{E}(R_i'Y_i)$, $\mathbb{E}((R_i'R_i)^{-1}R_i'Y_i)$, and $\mathbb{E}(Y_iR_i(R_i'R_i)^{-1}R_i'Y_i)$ are finite.

The general result in Theorem 2 presented in the next section provides insights on what type of information is used to construct the bounds in Theorem 1. It can be shown that, under the additional regularity conditions stated as Assumption 8 in the next section, the bounds in Theorem 1 are the sharp bounds of μ_e when the conditional moment restriction (2) is replaced by the following unconditional moment restrictions:

$$\mathbb{E}\left(\sum_{t=1}^{T}(R_{it}'B_i)\varepsilon_{it}\right) = 0, \quad \text{and} \quad \mathbb{E}\left(\sum_{t=1}^{T}R_{it}\varepsilon_{it}\right) = 0, \tag{10}$$

where the first restriction is interpreted as that the "error term" (ε_{it}) is orthogonal to the "explained term" ($R_{it}'B_i$), and the second is interpreted as that ε_{it} is orthogonal to the current-period regressors R_{it}, on average across t. For empirical applications, the amount of information contained in (10) is small relative to that in (2), and its refinement will be discussed later in this subsection. From a theoretical perspective, Theorem 1 suggests that the two moment conditions in (10) are the key identifying restrictions that yield finite L and U, out of the infinite number of unconditional moment restrictions in (9) that is equivalent to (2).

I now explain the intuition behind Theorem 1, focusing on the upper bound U. Consider a Lagrangian where the objective function is the parameter of interest $e'B_i$ and the constraints are the moment functions in (10):

$$Q(\lambda, \mu, W_i, B_i) = e'B_i + \lambda\sum_{t=1}^{T}(R_{it}'B_i)\varepsilon_{it} + \mu'\sum_{t=1}^{T}R_{it}\varepsilon_{it},$$

where $\lambda \in \mathbb{R}$ and μ has the same dimension as R_{it}. Note that $\mathbb{E}(Q) = \mathbb{E}(e'B_i) = \mu_e$ because the constraints have zero expectations by (10).

If I substitute $\varepsilon_{it} = Y_{it} - R'_{it}B_i$ into Q and use the matrix notations R_i and Y_i, I obtain the expression:

$$Q(\lambda, \mu, W_i, B_i) = e'B_i + \lambda Y_i' R_i B_i - \lambda B_i'(R_i'R_i)B_i + \mu' R_i' Y_i - \mu' R_i' R_i B_i.$$

This is a quadratic polynomial in B_i whose second-order derivative is

$$\frac{d^2 Q}{dB_i dB_i'} = -2\lambda(R_i'R_i).$$

If $\lambda > 0$, then this second-order derivative is a negative definite matrix, in which case Q attains a global maximum at the solution to the first-order condition $dQ/dB_i = 0$. Let $P = \max_{b \in \mathbb{R}^{q+p}} Q(\lambda, \mu, W_i, b)$ be the resulting maximum, which is only a function of (λ, μ, W_i) since B_i is "maximized out." Then, by construction:

$$P(\lambda, \mu, W_i) = \max_{b \in \mathbb{R}^{q+p}} Q(\lambda, \mu, W_i, b) \geq Q(\lambda, \mu, W_i, B_i),$$

which implies

$$\mathbb{E}(P(\lambda, \mu, W_i)) \geq \mathbb{E}(Q(\lambda, \mu, W_i, B_i)) = \mu_e.$$

This shows that $\mathbb{E}(P)$ is an upper bound of μ_e for any choice of $\lambda > 0$ and μ. I then obtain a smallest upper bound for μ_e by minimizing $\mathbb{E}(P)$ with respect to $\lambda > 0$ and μ:

$$\min_{\lambda > 0,\ \mu} \mathbb{E}(P(\lambda, \mu, W_i)) \geq \mu_e.$$

This coincides with U in Theorem 1. The lower bound can be obtained by repeating the same process with $\lambda < 0$.

As discussed earlier, the amount of information used to construct the bounds in Theorem 1, namely the moment restrictions in (10), is small relative to that in (2). I now develop a refinement of Theorem 1.

For each t, choose a vector of observable random variables S_{it} such that $\mathbb{E}(S_{it}\varepsilon_{it}) = 0$ under (2). For example, one may choose S_{it} to be $S_{it} = R_{it}$ (the vector of current regressors) or $S_{it} = (Z_i', X_i^{t\prime})'$ (the vector of the full history of Z_{it} and the current history of X_{it}). One may also choose S_{it} to include the square terms such as X_{it}^2 and Z_{it}^2. The dimension of S_{it} is allowed to vary across t. Consider the following assumption:

Assumption 3. $\mathbb{E}(S_{it}S_{it}')$ is positive definite for each $t = 1, \ldots, T$.

Assumption 3 states that the variance matrix of S_{it} has full rank. It is implied by Assumption 2 if $S_{it} = R_{it}$. Assumption 3 is trivially violated if S_{it} includes duplicate variables, for example, if $S_{it} = (X'_{it}, X'_{it}, Z'_{it})'$. However, it is not violated if S_{it} and S_{iv} for $t \neq v$ have duplicate variables.

I now state a refinement of Theorem 1 under Assumptions 1 to 3. For brevity of notation, define a block diagonal matrix

$$S_i \equiv \begin{pmatrix} S_{i1} & 0 & \cdots & 0 \\ 0 & S_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & S_{iT} \end{pmatrix}$$

where S_{it} appears in the diagonal as a column vector, so that S_i has T columns. In addition, define

$$\begin{aligned} \mathcal{V}_S &= \mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' S_i'), \\ \mathcal{Y}_S &= \mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' Y_i), \\ \mathcal{P}_S &= \mathbb{E}(S_i R_i (R_i' R_i)^{-1}), \\ Y_S &= \mathbb{E}(S_i Y_i), \\ m_0 &= \mathbb{E}(Y_i' R_i (R_i' R_i)^{-1} R_i' Y_i). \end{aligned}$$

Proposition 2. *Suppose that Assumptions 1 to 3 hold. Then $\mathcal{V}_S$ is invertible, and $L_S \leq \mu_e \leq U_S$ where*

$$[L_S, U_S] = \left[\mathcal{B}_S - \frac{1}{2}\sqrt{\mathcal{E}_S \mathcal{D}_S},\ \ \mathcal{B}_S + \frac{1}{2}\sqrt{\mathcal{E}_S \mathcal{D}_S} \right]$$

and

$$\begin{aligned} \mathcal{B}_S &= \frac{1}{2} e' \mathbb{E}(\widehat{B}_i) + \frac{1}{2} e' \mathcal{P}_S' \mathcal{V}_S^{-1} (2Y_S - \mathcal{Y}_S), \\ \mathcal{E}_S &= e' \mathbb{E}((R_i' R_i)^{-1}) e - e' \mathcal{P}_S' \mathcal{V}_S^{-1} \mathcal{P}_S e, \\ \mathcal{D}_S &= m_0 - (2Y_S - \mathcal{Y}_S)' \mathcal{V}_S^{-1} (2Y_S - \mathcal{Y}_S). \end{aligned}$$

Similarly to Theorem 1, under the additional regularity conditions stated as Assumption 8 in the next section, it can be shown that the bounds in Proposition 2 are the sharp bounds of μ_e if (2) is replaced by the following unconditional moments:

$$\mathbb{E}\left(\sum_{t=1}^{T} (R_{it}' B_i) \varepsilon_{it} \right) = 0, \quad \text{and} \quad \mathbb{E}\left(S_{it} \varepsilon_{it} \right) = 0 \quad \text{for} \quad t = 1, \ldots, T, \tag{11}$$

where the first expression gives one unconditional moment restriction, and the second expression gives $\dim(S_{it})$ unconditional moment restrictions for each t.

While (11) still contains less information than (2), it is found to be sufficiently informative in practice. In a numerical illustration in Section 3.4, I compare the outer bounds in Proposition 2 with the sharp bounds of μ_e calculated using the general result in Online Appendix B.3, finding that the two bounds are not excessively apart. Moreover, the empirical application in Section 6 shows that Proposition 2 can produce highly informative bounds. In addition, the closed-form expressions in Proposition 2 lead to a simple estimation and inference procedure that is robust to overidentification and model misspecification, which are generally not simple to deal with in partially identified models.

3.3 Extension to models with homogeneous coefficients

In this subsection, I extend the partial identification results of the previous subsection to the model that also involves homogeneous coefficients. Recall the expressions for the model introduced in (5) and (6):

$$Y_{it} = R_{it}'B_i + M_{it}'\delta + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it}|B_i, Z_i, X_i^t, \mathsf{Z}_i, \mathsf{X}_i^t) = 0, \qquad t = 1, \ldots, T,$$

where $M_{it} = (\mathsf{Z}_{it}', \mathsf{X}_{it}')'$ denotes the regressors with homogeneous coefficients. Let $U_{it} = (R_{it}', M_{it}')'$ be the vector of all regressors, and let M_i and U_i be random matrices stacking M_{it}' and U_{it}' rowwise across t, hence having T rows, respectively.

Similarly to the last subsection, choose a vector of observable random variables S_{it} such that $\mathbb{E}(S_{it}\varepsilon_{it}) = 0$ under (6). Consider the following modifications to Assumptions 1 and 2 and the restatement of Assumption 3.

Assumption 4. (Y_i, R_i, M_i, B_i) and δ satisfy (5) and (6).

Assumption 5. $R_i'R_i$ is positive definite with probability 1, and $\mathbb{E}(U_i'U_i)$ is positive definite.

Assumption 6. $\mathbb{E}(S_{it}S_{it}')$ is positive definite for each $t = 1, \ldots, T$.

Note that Assumption 5 only requires that $\mathbb{E}(M_i'M_i)$ is positive definite, rather than $M_i'M_i$ itself. Therefore, the no-multicollinearity requirement for M_{it} is the same as those for the regressors in standard fixed effect models.

Under these assumptions, the following proposition extends the bounds in Proposition 2 to the model defined in (5) and (6). For brevity of notation, define

$$
\begin{aligned}
\mathcal{V}_M &= \mathbb{E}(M_i' R_i (R_i' R_i)^{-1} R_i' M_i), & \mathcal{C} &= \mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' M_i), \\
\mathcal{Y}_M &= \mathbb{E}(M_i' R_i (R_i' R_i)^{-1} R_i' Y_i), & C &= \mathbb{E}(S_i M_i), \\
\mathcal{P}_M &= \mathbb{E}(M_i' R_i (R_i' R_i)^{-1}), & M_0 &= \mathbb{E}(M_i' M_i). \\
Y_M &= \mathbb{E}(M_i' Y_i),
\end{aligned}
$$

Proposition 3. *Suppose that Assumptions 4 to 6 hold. Then both the matrix $\mathcal{V}_M - M_0$ and the matrix*

$$
\mathcal{V} = \mathcal{V}_S - (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1}(C - \mathcal{C})'
$$

are invertible, and $L_M \leq \mu_e \leq U_M$ where

$$
[L_M, U_M] = \left[\mathcal{B}_M - \frac{1}{2}\sqrt{\mathcal{E}_M \mathcal{D}_M},\ \mathcal{B}_M + \frac{1}{2}\sqrt{\mathcal{E}_M \mathcal{D}_M}\right]
$$

and

$$
\begin{aligned}
\mathcal{B}_M &= \frac{1}{2} e' \mathbb{E}(\widehat{B}_i) + \frac{1}{2}(\mathcal{P}_S e + (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1} \mathcal{P}_M e)' \mathcal{V}^{-1} \times \\
&\qquad\qquad (2Y_S - \mathcal{Y}_S - (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1}(Y_M - \mathcal{Y}_M)), \\
\mathcal{E}_M &= e' \mathbb{E}((R_i' R_i)^{-1}) e \\
&\quad - (\mathcal{P}_S e + (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1} \mathcal{P}_M e)' \mathcal{V}^{-1} (\mathcal{P}_S e + (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1} \mathcal{P}_M e), \\
\mathcal{D}_M &= m_0 - (2Y_S - \mathcal{Y}_S - (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1}(Y_M - \mathcal{Y}_M))' \mathcal{V}^{-1} \times \\
&\qquad\qquad (2Y_S - \mathcal{Y}_S - (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1}(Y_M - \mathcal{Y}_M)).
\end{aligned}
$$

The empirical application in Section 6 shows that Proposition 3 can produce highly informative bounds. The empirical application involves a total of 59 regressors in M_{it} (and 58 in an alternative specification), demonstrating the practicality of Proposition 3 even when the number of regressors with homogeneous coefficients is large.

3.4 Numerical illustration

This subsection provides a numerical illustration of the sizes of the identified sets presented in the previous sections in a simple panel data model. This highlights the practical implications of considering unconditional moment restrictions instead of the conditional ones. Specifically, consider the model

$$
Y_{it} = \gamma_i + \beta_i X_{it} + \varepsilon_{it}, \qquad t = 1, \ldots, T,
$$

where

$$\mathbb{E}(\varepsilon_{it}|\gamma_i, \beta_i, X_i^t) = 0, \qquad t = 1, \dots, T. \tag{12}$$

For this model, I numerically compute the sharp identified set of $\mathbb{E}(\beta_i)$ under the conditional moment restriction in (12), and compare it to the outer identified set in Proposition 2, which are based on the unconditional moment restrictions in (11). Computation of the sharp identified set of $\mathbb{E}(\beta_i)$ is generally prohibitively expensive (see the discussion in Section 4), but it becomes relatively tractable when (γ_i, β_i, X_i) has a small number of discrete support points, where the sharp characterization reduces to solving optimization problems over a large but finite-dimensional Euclidean spaces.

Let γ_i and β_i be independent discrete random variables such that $\gamma_i \in \{-1, 0, 1\}$ with equal probabilities and $\beta_i \in \{0, 0.5, 1\}$ with equal probabilities. In addition, let ε_{it} be independent of (γ_i, β_i) and $\varepsilon_{it} \in \{-1, 0, 1\}$ with equal probabilities. Lastly, let $X_{i1} = 1$, and define for $t \geq 2$:

$$X_{it} = \begin{cases} -1 & \text{if } Y_{i,t-1} < -1, \\ 0 & \text{if } -1 \leq Y_{i,t-1} < 1, \\ 1 & \text{if } Y_{i,t-1} \geq 1, \end{cases}$$

so that X_{it} depends on $Y_{i,t-1}$. Under this data generating process, I compute both the sharp and the outer bounds of $\mathbb{E}(\beta_i)$ for $T \in \{3, 4, 5\}$. I also calculate the outer bounds for $T \in \{6, 8\}$ to illustrate how the outer bound tightens as T increases. I choose $S_{it} = (1, X_{i1}, \dots, X_{it})'$ to compute the outer bounds in Proposition 2.

The calculated sharp and outer bounds are presented in Table 1. Note that the data generating process implies $\mathbb{E}(\beta_i) = 0.5$. Although the outer bounds are wider than the sharp bounds, the gap is modest, and the outer bounds remain sufficiently informative to concentrate around $\mathbb{E}(\beta_i) = 0.5$.

	$T = 3$	$T = 4$	$T = 5$	$T = 6$	$T = 8$
Sharp	[0.401, 0.593]	[0.452, 0.552]	[0.473, 0.532]	-	-
Outer	[0.216, 0.617]	[0.267, 0.613]	[0.306, 0.613]	[0.330, 0.613]	[0.368, 0.598]

Table 1: Numerical illustration of the sharp and the outer identified sets. Sharp refers to the sharp identified set of $\mathbb{E}(\beta_i)$ under (12), and Outer refers to the outer bounds of $\mathbb{E}(\beta_i)$ under (11). The sharp identified sets for $T = 6$ and $T = 8$ are not computed because they are computationally prohibitive.

4 Identification of the general parameters

This section presents a general partial identification result for dynamic random coefficient models. This section is structured into two subsections. First, I present a general partial identification result for a generic parameter. Second, I apply this general result to derive the bounds for the variance and the CDF of the random coefficient distribution.

4.1 Identification of the general parameters

Recall that $W_i \in \mathcal{W}$ is the vector of observable variables and $B_i \in \mathcal{B}$ is the vector of unobservable random coefficients. I consider parameters of the form

$$\theta = \mathbb{E}(m(W_i, B_i))$$

where the function $m : \mathcal{W} \times \mathcal{B} \mapsto \mathbb{R}$ is known. I consider a generic set of unconditional moment restrictions:

Assumption 7. The random vectors (W_i, B_i) satisfy:

$$\mathbb{E}(\phi_k(W_i, B_i)) = 0, \quad k = 1, \ldots, K,$$

where $\phi_k : \mathcal{W} \times \mathcal{B} \mapsto \mathbb{R}$ are known moment functions and $K \in \mathbb{N}$ is the number of moment restrictions.

Note that, in the asymptotics, K is fixed when $N \to \infty$. Note also that ε_{it} does not appear in Assumption 7 because ε_{it} is determined by (W_i, B_i) via the relationship $\varepsilon_{it} = Y_{it} - R_{it}' B_i$. More generally, without connection to random coefficient models, Assumption 7 imposes generic unconditional moment restrictions that involve both observed and unobserved random vectors. A more general formulation that also involves conditional moment restrictions is studied in Online Appendix B.3.

I characterize the sharp identified set of θ under Assumption 7 and the regularity conditions that are introduced below. To do so, I first recast the identification problem as a linear programming problem. I then show that its dual representation yields a tractable characterization of the identified set.

Let $P \in \mathcal{M}_{W \times B}$, where $\mathcal{M}_{W \times B}$ is the linear space of finite and countably additive signed Borel measures on $\mathcal{W} \times \mathcal{B}$, equipped with the total variation norm. Let $P_W \in \mathcal{M}_W$

be the observed marginal distribution of W_i. The sharp identified set I of θ is *defined* by:

$$I \equiv \left\{ \int m(w,b)dP \;\middle|\; \begin{aligned} &P \in \mathcal{M}_{W\times B}, \quad P \geq 0, \quad \int dP = 1, \\ &\int \phi_k(w,b)dP = 0, \quad k = 1,\ldots,K, \\ &\int P(w,db) = P_W(w) \text{ for all } w \in \mathcal{W} \end{aligned} \right\}.$$

The set I is the collection of all $\int m(W_i, B_i)dP$ values over P such that (i) P is a probability distribution of (W_i, B_i), (ii) P satisfies the moment restrictions, and (iii) the marginal distribution of W_i implied from P equals the observed distribution P_W. Dependence of I on m, P_W, and the ϕ_ks are suppressed in the notation.

All defining properties of I are linear in P, which means that I is a convex set in $\mathbb{R}$ (i.e., an interval). Therefore, I can be characterized by its lower and upper bounds. The sharp lower bound L of I is *defined* by:

$$\min_{P\in\mathcal{M}_{W\times B},\, P\geq 0} \int m(w,b)dP \qquad \text{subject to} \qquad \begin{aligned} &\int \phi_k(w,b)dP = 0, \quad k = 1,\ldots,K, \\ &\int P(w,db) = P_W(w) \text{ for all } w \in \mathcal{W}. \end{aligned} \tag{13}$$

Note that the constraint $\int dP = 1$ is omitted in (13), because it is implied by the constraint $\int P(w,db) = P_W(w)$ where P_W is a probability distribution.

Equation (13) is a linear program in P, with the caveat that P is an infinite-dimensional object. It is not a tractable characterization of L for dynamic random coefficient models, in the sense that the estimation methods it imply are computationally infeasible. For example, discretizing the space of (W_i, B_i) and solving the discretized problem (Honoré and Tamer, 2006; Gunsilius, 2019) is computationally infeasible because the dimension of (W_i, B_i) is large. Recall that W_i contains the full history of regressors and dependent variables and B_i contains all random coefficients. For the random coefficient model with R regressors and T waves, P is a distribution on an $(RT + T + R)$-dimensional space.

My approach is to use the dual representation of (13) obtained by the duality theorem for infinite-dimensional linear programming (Galichon and Henry, 2009; Schennach, 2014). I consider the following regularity conditions:

Assumption 8. The following conditions hold.

(i) $\mathcal{W} \times \mathcal{B}$ is a compact set in a Euclidean space.

(ii) $(m, \phi_1, \ldots, \phi_K)$ are bounded Borel measurable functions on $\mathcal{W} \times \mathcal{B}$.

Under these conditions, the following theorem characterizes the sharp identified set of θ using the dual representation of (13) and the corresponding problem for the upper bound.

Theorem 2. *Suppose Assumptions 7 and 8 hold. Let $\lambda = (\lambda_1, \ldots, \lambda_K)' \in \mathbb{R}^K$. Then $I = [L, U]$ where*

$$L = \max_{\lambda \in \mathbb{R}^K} \mathbb{E}\left[\min_{b \in \mathcal{B}} \left\{ m(W_i, b) + \sum_{k=1}^{K} \lambda_k \phi_k(W_i, b) \right\}\right] \tag{14}$$

and

$$U = \min_{\lambda \in \mathbb{R}^K} \mathbb{E}\left[\max_{b \in \mathcal{B}} \left\{ m(W_i, b) + \sum_{k=1}^{K} \lambda_k \phi_k(W_i, b) \right\}\right] \tag{15}$$

provided that the optimization problems in (14) and (15) possess finite solutions.

Note that the result in Theorem 2 is not specific to dynamic random coefficient models. It is a general duality result for moment equality models where the moment functions involve both observables and unobservables (Schennach, 2014; Li, 2018).

The characterization that also involves conditional moment restrictions is developed in Online Appendix B.3. To illustrate, consider the following conditional moment restrictions:

$$\mathbb{E}(\psi_k(W_i, B_i) | \mathsf{W}_{ik}, \mathsf{B}_{ik}) = 0, \qquad k = 1, \ldots, K,$$

where W_{ik} and B_{ik} are subvectors of W_i and B_i. Note that (2) has T moment restrictions of this type, one for each $t = 1, \ldots, T$. Assume that W_i and B_i are absolutely continuous with respect to the Lebesgue measure, and that the regularity conditions stated as Assumption 13 (i)-(iii) in Online Appendix B.3 hold. Then, under these assumptions, Theorem 3 in Online Appendix B.3 implies that the sharp lower bound of θ is given by

$$L = \max_{\{\mu_k(\mathsf{w}_k, \mathsf{b}_k) \in L^2(\mathsf{W}_{ik}, \mathsf{B}_{ik})\}_{k=1}^{K}} \mathbb{E}\left[\min_{b \in \mathcal{B}} \left\{ m(W_i, b) + \sum_{k=1}^{K} \mu_k(\mathsf{W}_{ik}, \mathsf{b}_k) \psi_k(W_i, b) \right\}\right], \tag{16}$$

where b_k is the subvector of b corresponding to B_{ik} and $\mu_k(\mathsf{w}_k, \mathsf{b}_k)$ is a square integrable function of $(\mathsf{W}_{ik}, \mathsf{B}_{ik})$, denoted by $L^2(\mathsf{W}_{ik}, \mathsf{B}_{ik})$. Therefore, for conditional moment restrictions, the dual representation involves optimization over the functional choice variables $\{\mu_k(\mathsf{w}_k, \mathsf{b}_k)\}_{k=1}^{K}$. In general, such functional optimization is not computationally tractable because the inner optimization problem over b is potentially highly nonconvex and (W_i, B_i) is potentially high-dimensional. For the random coefficient model with R regressors and T waves, each μ_k is a function on a space of dimension at most $(RT + R)$, and one must optimize over K such functions in (16). In contrast, (14) involves optimization over the finite-dimensional Euclidean space $\mathbb{R}^K$. In the previous subsection, I used a parsimonious set of moment restrictions in (11) to derive closed-form bounds for the mean parameters that are computationally efficient. In the next subsection, using the same parsimonious set, I derive computationally efficient bounds of the variance and the CDF parameters. While these bounds are the outer bounds relative to the sharp bounds in (16), I demonstrate in the numerical illustration in Section 3.4 and the empirical application in Section 6 that they produce informative bounds in practice.

The condition that (14) and (15) possess finite solutions is mild due to the following key property. Define the value functions of the inner optimization problems in (14) and (15) as G_L and G_U, respectively:

$$G_L(\lambda, w) = \min_{b \in \mathcal{B}} \left\{ m(w, b) + \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right\},$$
$$G_U(\lambda, w) = \max_{b \in \mathcal{B}} \left\{ m(w, b) + \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right\}.$$

Note that, given the model ingredients m and $\phi_1, \ldots, \phi_K$, these are deterministic functions of (λ, w). They have the following key property.

Proposition 4. *$G_L(\lambda, w)$ is globally concave in λ for every w, and $G_U(\lambda, w)$ is globally convex in λ for every w.*

Since concave and convex functions on $\mathbb{R}^K$ are continuous, (14) and (15) possess finite solutions whenever the optimizers in λ lie in the interior of $\mathbb{R}^K$. For dynamic random coefficient models in (1) and (2), this can be achieved by a suitable choice of the moment functions $(\phi_1, \ldots, \phi_K)$ derived from (2), which I illustrate in the next subsection for the variance and the CDF parameters.

Using the definitions of G_L and G_U, the bounds in Theorem 2 can be written as

$$L = \max_{\lambda \in \mathbb{R}^K} \mathbb{E}\left[G_L(\lambda, W_i)\right], \quad \text{and} \quad U = \min_{\lambda \in \mathbb{R}^K} \mathbb{E}\left[G_U(\lambda, W_i)\right].$$

Under suitable conditions, G_L and G_U are differentiable when $K = 1$ (Milgrom and Segal, 2002, Theorem 3), which can be extended to show that G_L and G_U are directionally differentiable for $K > 1$. Proposition 4 then implies that the optimization problems over λ can be solved using fast convex optimization algorithms such as gradient descent, provided that the inner optimization problems over b can be solved efficiently. In the next subsection, I illustrate the choice of the moment functions for the variance and the CDF parameters that admits computationally efficient solutions to the inner optimization problems.

A direct consequence of Theorem 2 is that θ is point-identified if and only if $L = U$.

Proof of Theorem 2 then implies a necessary and sufficient condition for point identification of θ, which I state as a separate lemma below.

Lemma 2. *Suppose that the assumptions of Theorem 2 hold. Suppose also that* (W_i, B_i) *are absolutely continuous with respect to the Lebesgue measure, and that their joint density is strictly positive on* $\mathcal{W} \times \mathcal{B}$*. Then* θ *is point-identified if and only if there exists a function* S^**, which is a linear functional on* $\mathcal{M}_W$*, and real numbers* $\lambda_1^*, \dots, \lambda_K^* \in \mathbb{R}$ *such that:*

$$m(W_i, B_i) + \sum_{k=1}^{K} \lambda_k^* \phi_k(W_i, B_i) = S^*(W_i)$$

almost surely on $\mathcal{W} \times \mathcal{B}$*. When such* S^* *exists,* θ *is identified by* $\theta = \mathbb{E}(S^*(W_i))$*.*

Lemma 2 states that θ is point-identified if and only if the Lagrangian reduces to a function of data only. Note that S^* can be considered as an unbiased estimator because the term $\sum_{k=1}^{K} \lambda_k^* \phi_k(W_i, B_i)$ has zero expectation.

Lastly, I highlight the connection between Theorem 2 and the support function approach of Beresteanu, Molchanov, and Molinari (2011). Let δ be a structural parameter and consider the moment conditions

$$\mathbb{E}(\phi_k(W_i, B_i, \delta)) = 0, \quad k = 1, \dots, K.$$

In what follows, I fix the value of δ and consider each $\phi_k(\cdot, \cdot, \delta)$ as a function of (W_i, B_i) only. In addition, I set $m(W_i, B_i) = 0$, so that $\theta = \mathbb{E}(m(W_i, B_i)) = 0$. In this case, the sharp lower bound of $\theta = 0$ is obtained by specializing (13) with $m = 0$:

$$L_{primal}(\delta) = \min_{P \in \mathcal{M}_{W \times B},\, P \geq 0} 0 \qquad \text{subject to} \qquad \begin{aligned} &\int \phi_k(w, b, \delta) dP = 0, \quad k = 1, \dots, K, \\ &\int P(w, db) = P_W(w) \ \text{ for all } w \in \mathcal{W}. \end{aligned}$$

The solution of this problem is trivially 0, but only if there exists a probability distribution P that satisfies the moment conditions. If no such P exists, the problem is infeasible, and I set $L_{primal}(\delta) = \infty$. This characterization is similar in spirit to those in Honoré and Tamer (2006), Honoré and Lleras-Muney (2006) and Molinari (2008), extended here to allow P to be a continuous distribution. I can then write the identified set of δ as

$$\{\delta \mid L_{primal}(\delta) = 0\}.$$

Theorem 2 then implies that the dual representation of $L_{primal}(\delta)$ is

$$L_{dual}(\delta) = \max_{\lambda \in \mathbb{R}^K} \mathbb{E} \left[\min_{b \in \mathcal{B}} \left\{ \sum_{k=1}^{K} \lambda_k \phi_k(W_i, b, \delta) \right\} \right]. \tag{17}$$

Using this, the identified set of δ can also be written as $\{\delta \mid L_{dual}(\delta) = 0\}$. This characterization coincides with the support function characterization in Beresteanu, Molchanov, and Molinari (2011, Section 4) given for regression coefficients with interval data. In particular, in their Theorem 4.1, the negative of their support function coincides with the inner objective function $\sum_{k=1}^{K} \lambda_k \phi_k(W_i, b, \delta)$ in (17), and their variable u coincides with the Lagrange multiplier λ in (17).

4.2 Examples: the variance and the CDF of random coefficients

In this subsection, I apply Theorem 2 to derive the bounds for the variance and the CDF parameters. Recall the dynamic random coefficient model defined in (1) and (2):

$$Y_{it} = R_{it}' B_i + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it} | B_i, Z_i, X_i^t) = 0, \qquad t = 1, \dots, T,$$

where $R_{it} = (Z_{it}', X_{it}')'$. I first consider the second moments of the random coefficients:

$$V_e = \mathbb{E}(e_1' B_i B_i' e_2) = \mathbb{E}(B_i' e_1 e_2' B_i),$$

where e_1 and e_2 are real-valued constant vectors chosen by the econometrician.

A full characterization of the bounds of V_e for generic choices of e_1 and e_2 is discussed in Online Appendix Section B.4. Here, I focus on the special case where $e_1 = e_2$ and this common vector has a single entry equal to 1 and zeros elsewhere. In this case, V_e is the second moment of a particular coefficient — a key ingredient of the variance parameter. This particular case deserves separate discussion because its bounds can be computed more efficiently than those in the general case.

In the case where $e_1 = e_2$ and this common vector has a single entry equal to 1 and zeros elsewhere, define $e_0 = e_1 e_1'$. Then e_0 is a diagonal matrix which has 1 in only one entry and zeros elsewhere. For example, if $B_i = (\beta_{i1}, \beta_{i2})'$ and $e_1 = e_2 = (0, 1)'$, then

$$e_0 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad V_e = \mathbb{E}(B_i' e_0 B_i) = \mathbb{E}(\beta_{i2}^2).$$

Recall the moment restrictions used for the bounds in Proposition 2, namely those in (11):

$$\mathbb{E}\left(\sum_{t=1}^{T}(R_{it}'B_i)\varepsilon_{it}\right)=0,\quad \text{and} \quad \mathbb{E}\left(S_{it}\varepsilon_{it}\right)=0 \quad \text{for} \quad t=1,\ldots,T.$$

Let $L=\sum_{t=1}^{T}\dim(S_{it})$. Applying Theorem 2 with these restrictions yields the following lower bound for V_e, denoted by L_V:

$$\begin{aligned}L_V &= \max_{\lambda\in\mathbb{R},\ \mu\in\mathbb{R}^L}\mathbb{E}\left[\min_{b\in\mathcal{B}}\left\{b'e_0b+\lambda b'R_i'(Y_i-R_ib)+\mu'S_i(Y_i-R_ib)\right\}\right]\\ &= \max_{\lambda\in\mathbb{R},\ \mu\in\mathbb{R}^L}\mathbb{E}\left[\min_{b\in\mathcal{B}}\left\{\mu'S_iY_i+(\lambda R_i'Y_i-R_i'S_i'\mu)'b-b'(\lambda R_i'R_i-e_0)b\right\}\right],\end{aligned}$$

Suppose that Assumptions 1 to 3 hold. Note that, in L_V, the objective function of the inner minimization problem is a quadratic polynomial in b, where the leading coefficient matrix is $-(\lambda R_i'R_i-e_0)=e_0-\lambda R_i'R_i$. Since e_0 is positive semidefinite, and since $R_i'R_i$ is positive definite by Assumption 2, this matrix is positive definite if $\lambda<0$, in which case the quadratic polynomial attains a finite closed-form minimum at the solution to the first order condition. On the other hand, if $\lambda\geq 0$, the polynomial's minimum diverges to $-\infty$ with positive probability, provided (Y_i,R_i,S_i) is non-degenerate. Consequently, a finite lower bound is obtained only when $\lambda<0$. I then obtain the following expression for the lower bound, which is a direct application of Theorem 2 and is stated without proof.

Proposition 5. *Suppose that Assumptions 1 to 3 and 8 hold. Then $L_V\leq V_e$ where*

$$L_V=\max_{\lambda<0,\ \mu\in\mathbb{R}^L}\mathbb{E}\left[\min_{b\in\mathcal{B}}\left\{\mu'S_iY_i+(\lambda R_i'Y_i-R_i'S_i'\mu)'b-b'(\lambda R_i'R_i-e_0)b\right\}\right].$$

Note that the objective function in Proposition 5 is concave in (λ,μ) by Proposition 4. Therefore, the maximization over (λ,μ) can be performed efficiently using standard convex optimization methods. The inner minimization over b can also be solved efficiently using quadratic optimization softwares.

An upper bound of V_e can be obtained similarly, but with a stronger assumption. Specifically, by applying Theorem 2, I obtain the following upper bound of V_e:

$$U_V=\min_{\lambda\in\mathbb{R},\ \mu\in\mathbb{R}^L}\mathbb{E}\left[\max_{b\in\mathcal{B}}\left\{\mu'S_iY_i+(\lambda R_i'Y_i-R_i'S_i'\mu)'b-b'(\lambda R_i'R_i-e_0)b\right\}\right].$$

The inner objective function is a quadratic polynomial in b, where the leading coefficient matrix is $-(\lambda R_i'R_i-e_0)=e_0-\lambda R_i'R_i$. If this matrix is negative definite, the quadratic polynomial attains a finite closed-form maximum at the solution to the first order condition, and otherwise the inner maximization problem diverges to $+\infty$ with positive probability. Therefore, a finite upper bound is obtained only in the region where $e_0 - \lambda R_i' R_i$ is negative definite, i.e., all of its eigenvalues are negative. Note first that this requires $\lambda > 0$, as otherwise $e_0 - \lambda R_i' R_i$ is positive semidefinite. Then, since e_0 is a diagonal matrix which has a single entry equal to 1 and zeros elsewhere, Weyl's inequality implies that the largest eigenvalue of $e_0 - \lambda R_i' R_i$ is bounded above by $1 - \lambda \nu$, where $\nu > 0$ is the smallest eigenvalue of $R_i' R_i$. Therefore, all eigenvalues of $e_0 - \lambda R_i' R_i$ are negative if $1 - \lambda \nu < 0$, which is equivalent to $\lambda > 1/\nu$. Under this additional condition, I obtain the following upper bound of V_e, which is a direct application of Theorem 2 and is stated without proof.

Assumption 9. There exists $\lambda_{min} > 0$ such that the smallest eigenvalue of $R_i' R_i$ is strictly larger than λ_{min} almost surely.

Proposition 6. *Suppose that Assumptions 1 to 3, 8 and 9 hold. Then $V_e \leq U_V$ where*

$$U_V = \min_{\lambda \geq \lambda_{min},\ \mu \in \mathbb{R}^L} \mathbb{E}\left[\max_{b \in \mathcal{B}} \left\{\mu' S_i Y_i + (\lambda R_i' Y_i - R_i' S_i' \mu)' b - b'(\lambda R_i' R_i - e_0) b\right\}\right].$$

Note that Assumption 9 is stronger than Assumption 2. While Assumption 2 requires the smallest eigenvalue of $R_i' R_i$ to be positive almost surely, Assumption 9 further requires that it is strictly bounded away from zero. I also derive the bounds for V_e that do not require Assumption 9 in Online Appendix B.4, but their estimation will involve computationally more intensive methods.

Next, I consider the CDF of the random coefficients. Consider the parameter of the form

$$F_{e,c} = \mathbb{P}(e' B_i \leq c) = \mathbb{E}(\mathbf{1}(e' B_i \leq c))$$

where e is a real-valued constant vector and c is a scalar. To derive the identified set of $F_{e,c}$, I consider the moment restrictions used for the bounds in Proposition 2, namely those in (11). Applying Theorem 2 with (11) yields the bounds with the inner objective function

$$\mathcal{L} = \mathbf{1}(e' B_i \leq c) + \lambda B_i' R_i' (Y_i - R_i B_i) + \mu' S_i (Y_i - R_i B_i)$$

that must be optimized over B_i for fixed (λ, μ). Note that the indicator $\mathbf{1}(e' B_i \leq c)$ partitions the support of B_i into two disjoint sets, where it equals to 1 on the set $\{B_i | e' B_i \leq c\}$ and 0 on the set $\{e' B_i > c\}$. Moreover, within each set, $\mathcal{L}$ reduces to a standard quadratic polynomial in B_i, which can be solved efficiently. Therefore, the optimization of $\mathcal{L}$ over B_i for fixed (λ, μ) can be carried out in two steps: (i) solve for the quadratic polynomial within each set, and then (ii) take the optimum between the two. I then obtain the following characterization for the identified set of $F_{e,c}$, which is a direct application of Theorem 2 and is stated without proof.

Proposition 7. *Suppose that Assumptions 1 to 3 and 8 hold. Then $L_F \leq F_{e,b} \leq U_F$ where*

$$L_F = \max_{\lambda<0,\ \mu\in\mathbb{R}^L} \mathbb{E}\left[G_{L,F}(W_i,\lambda,\mu)\right], \quad \textit{and} \quad U_F = \min_{\lambda>0,\ \mu\in\mathbb{R}^L} \mathbb{E}\left[G_{U,F}(W_i,\lambda,\mu)\right],$$

where

$$\begin{aligned} G_{L,F}(W_i,\lambda,\mu) = \min\Big\{ &\min_{b\ \in\ \{b\in\mathcal{B}\ \mid\ e'b\leq c\}} \Big[1+\lambda b'R_i'(Y_i-R_ib)+\mu'S_i(Y_i-R_ib)\Big], \\ &\min_{b\ \in\ \{b\in\mathcal{B}\ \mid\ e'b>c\}} \Big[\lambda b'R_i'(Y_i-R_ib)+\mu'S_i(Y_i-R_ib)\Big]\Big\}, \end{aligned}$$

and

$$\begin{aligned} G_{U,F}(W_i,\lambda,\mu) = \max\Big\{ &\max_{b\ \in\ \{b\in\mathcal{B}\ \mid\ e'b\leq c\}} 1+\lambda b'R_i'(Y_i-R_ib)+\mu'S_i(Y_i-R_ib), \\ &\max_{b\ \in\ \{b\in\mathcal{B}\ \mid\ e'b>c\}} \Big[\lambda b'R_i'(Y_i-R_ib)+\mu'S_i(Y_i-R_ib)\Big]\Big\}. \end{aligned}$$

If B_i is continuous, then one may replace the set $\{b \in \mathcal{B} \mid e'b > c\}$ with its closure $\{b \in \mathcal{B} \mid e'b \geq c\}$, which facilitates estimation and inference.

5 Estimation and inference

This section discusses estimation and inference for the identified sets derived in Sections 3 and 4. This section is structured into two subsections. In the first, I consider inference for the mean parameters discussed in Section 3. I exploit their simple closed-form expressions to present a procedure that is both straightforward to implement and robust to overidentification and model misspecification. In the second, I consider inference for the general parameters discussed in Section 4, presenting a procedure under the assumption of correct model specification.

5.1 Estimation and inference for the mean parameters

In this subsection, I discuss estimation and inference for the mean parameters, focusing on the refined bounds in Propositions 2 and 3. In what follows, I present a procedure for the bounds in Proposition 2. The same procedure applies to the bounds in Proposition 3.

Note that the bounds $[L_S, U_S]$ in Proposition 2 are deterministic functions of the following moments:

$$
\begin{aligned}
V_0 &= \mathcal{V}_S = \mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' S_i'), \\
Y_0 &= 2Y_S - \mathcal{Y}_S = \mathbb{E}(2S_i Y_i - S_i R_i (R_i' R_i)^{-1} R_i' Y_i), \\
P_0 &= \mathcal{P}_S e = \mathbb{E}(S_i R_i (R_i' R_i)^{-1} e), \\
m_0 &= \mathbb{E}(Y_i' R_i (R_i' R_i)^{-1} R_i' Y_i), \\
b_0 &= \mathbb{E}(e' \widehat{B}_i) = \mathbb{E}(e' (R_i' R_i)^{-1} R_i' Y_i), \\
R_0 &= \mathbb{E}((R_i' R_i)^{-1}).
\end{aligned}
$$

Let D_i be the vector that collects all of the entries inside these expectations. In other words, D_i is defined as

$$
\begin{aligned}
D_i = \Big(&\text{vech}(S_i R_i (R_i' R_i)^{-1} R_i' S_i')',\ (2S_i Y_i - S_i R_i (R_i' R_i)^{-1} R_i' Y_i)',\ (S_i R_i (R_i' R_i)^{-1} e)', \\
&Y_i' R_i (R_i' R_i)^{-1} R_i' Y_i,\ e' (R_i' R_i)^{-1} R_i' Y_i,\ \text{vech}((R_i' R_i)^{-1})'\Big)'.
\end{aligned}
$$

Note that $\mathbb{E}(D_i) = (\text{vech}(V_0)', Y_0', P_0', m_0, b_0, \text{vech}(R_0)')'$. Now, given an independent and identically distributed (i.i.d.) sample $\{D_i\}_{i=1}^N$ of size N, define

$$
\overline{D}_N = \frac{1}{N} \sum_{i=1}^N D_i.
$$

I assume that $\overline{D}_N$ is asymptotically normal with rate $\sqrt{N}$, which holds if the conditions for multivariate Central Limit Theorem hold for D_i (Van der Vaart, 2000, Section 2).

Assumption 10. $\sqrt{N}(\overline{D}_N - \mathbb{E}(D_i))$ converges in distribution to $N(0, V_D)$ for some variance matrix V_D.

Now I discuss estimation and inference for $[L_S, U_S]$ under assumptions of Proposition 2 and Assumption 10. Recall that the expressions for $[L_S, U_S]$ are:

$$
[L_S, U_S] = \left[\mathcal{B}_S - \frac{1}{2}\sqrt{\mathcal{E}_S \mathcal{D}_S},\ \mathcal{B}_S + \frac{1}{2}\sqrt{\mathcal{E}_S \mathcal{D}_S}\right].
$$

Let $\widehat{\mathcal{B}}_S$, $\widehat{\mathcal{E}}_S$, and $\widehat{\mathcal{D}}_S$ be the sample counterparts of $\mathcal{B}_S$, $\mathcal{E}_S$, and $\mathcal{D}_S$ calculated with $\overline{D}_N$. For example, $\widehat{\mathcal{B}}_S$ is given by

$$\widehat{\mathcal{B}}_S = \frac{1}{2}e'\left(\frac{1}{N}\sum_{i=1}^{N}(R_i'R_i)^{-1}R_i'Y_i\right) + \frac{1}{2}e'\left(\frac{1}{N}\sum_{i=1}^{N}(R_i'R_i)^{-1}R_i'S_i'\right) \times \\ \left(\frac{1}{N}\sum_{i=1}^{N}S_iR_i(R_i'R_i)^{-1}R_i'S_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}(2S_iY_i - S_iR_i(R_i'R_i)^{-1}R_i'Y_i)\right).$$

Then, define an estimator of $[L_S, U_S]$ as

$$[\widehat{L}_S, \widehat{U}_S] = \left[\widehat{\mathcal{B}}_S - \frac{1}{2}\sqrt{\widehat{\mathcal{E}}_S\widehat{\mathcal{D}}_S},\ \widehat{\mathcal{B}}_S + \frac{1}{2}\sqrt{\widehat{\mathcal{E}}_S\widehat{\mathcal{D}}_S}\right].$$

Since $[L_S, U_S]$ is a smooth function of $\mathbb{E}(D_i)$ provided that $\mathcal{E}_S > 0$ and $\mathcal{D}_S > 0$, the Delta method (Van der Vaart, 2000, Section 3) implies that $[\widehat{L}_S, \widehat{U}_S]$ is asymptotically normal. A key practical issue, however, is that the quantity $\widehat{\mathcal{D}}_S$ may be negative in finite samples, causing the term $\sqrt{\widehat{\mathcal{E}}_S\widehat{\mathcal{D}}_S}$ and thus the estimator $[\widehat{L}_S, \widehat{U}_S]$ to be not well-defined. This issue is related to a well-known challenge in inference for partially identified models — overidentification and model misspecification. In what follows, I discuss this issue in detail and propose an inference procedure that addresses it.

Recall that the bounds $[L_S, U_S]$ arise as the dual representations of the primal problem in (13) (and the corresponding problem for the upper bound) based on the moment restrictions in (11). It can be shown that the estimated bounds $[\widehat{L}_S, \widehat{U}_S]$ are the dual of the sample version of (13) where the population distribution P_W is replaced with the finite-sample empirical distribution $\hat{P}_W$. Overidentification then arises when the population problem (13) is feasible but its sample version with $\hat{P}_W$ is infeasible. This mirrors the familiar overidentification problem in generalized method of moments (GMM) estimation where, even if the population satisfies all the moment restrictions so that the population GMM criterion achieves zero, the finite sample may not satisfy all moment restrictions simultaneously, yielding a strictly positive sample GMM criterion[3]. In terms of the closed-form expressions of Proposition 2, this corresponds to having $\mathcal{D}_S > 0$ but $\widehat{\mathcal{D}}_S < 0$. In contrast, model misspecification arises when the population problem (13) itself is infeasible. In this case, the population quantity $\mathcal{D}_S$ is negative, and thus its sample counterpart $\widehat{\mathcal{D}}_S$ is also likely to be negative.

A recently growing literature on inference under misspecification in partially identified models (Stoye, 2020; Andrews and Kwon, 2024) propose solutions to these issues.

[3] For the empirical likelihood approach, this translates into the empirical likelihood criterion failing to attain its optimum at equal probabilities.

In what follows, I adopt the procedure of Stoye (2020) who provides a simple, easy-to-implement method for conducting inference on bounds that are smooth functions of the moments.

To apply the procedure of Stoye (2020), I first construct a smooth approximation and extension of the bounds $[L_S, U_S]$ that remains well-defined for any values of $\mathcal{B}_S$, $\mathcal{E}_S$, and $\mathcal{D}_S$. Let $r > 0$ be a small constant, and define the smoothed square root function

$$s(x,y) = \sqrt{\frac{xy + \sqrt{(xy)^2 + r^2}}{2}}.$$

For small $r > 0$, this function satisfies $s(x,y) = \sqrt{xy} + O(r)$ if $xy > 0$, and $s(x,y) = O(r)$ if $xy < 0$. In other words, $s(x,y)$ coincides with the ordinary square root when $xy > 0$ and vanishes if $xy < 0$. In addition, because of the term $r^2 > 0$, it is smooth everywhere, including at $xy = 0$. I then define the smooth approximation and extension of $[L_S, U_S]$ as:

$$[L_{Smth}, U_{Smth}] = \left[\mathcal{B}_S - \frac{1}{2}\left(s(\mathcal{E}_S, \mathcal{D}_S) - s(\mathcal{E}_S, -\mathcal{D}_S)\right),\ \mathcal{B}_S + \frac{1}{2}\left(s(\mathcal{E}_S, \mathcal{D}_S) - s(\mathcal{E}_S, -\mathcal{D}_S)\right)\right].$$

Note that $\mathcal{E}_S > 0$ even under overidentification and misspecification because

$$\begin{aligned}\mathcal{E}_S &= e'\mathbb{E}((R_i'R_i)^{-1})e - e'\mathcal{P}_S'\mathcal{V}_S^{-1}\mathcal{P}_S e\\ &= \mathbb{E}\left((e' - e'\mathbb{E}((R_i'R_i)^{-1}R_i'S_i')S_iR_i)(R_i'R_i)^{-1}(e - R_i'S_i'\mathbb{E}(S_iR_i(R_i'R_i)^{-1})e)\right),\end{aligned}$$

which is a quadratic form associated with a positive definite matrix. Therefore, $\mathcal{D}_S$ is the only quantity that can become negative in the square root function. Then, if $\mathcal{D}_S > 0$, $[L_{Smth}, U_{Smth}]$ simplifies to

$$[L_{Smth}, U_{Smth}] \approx \left[\mathcal{B}_S - \frac{1}{2}s(\mathcal{E}_S, \mathcal{D}_S),\ \mathcal{B}_S + \frac{1}{2}s(\mathcal{E}_S, \mathcal{D}_S)\right],$$

which coincides with $[L_S, U_S]$ up to the error term $O(r)$. If $\mathcal{D}_S < 0$, then

$$[L_{Smth}, U_{Smth}] \approx \left[\mathcal{B}_S + \frac{1}{2}s(\mathcal{E}_S, -\mathcal{D}_S),\ \mathcal{B}_S - \frac{1}{2}s(\mathcal{E}_S, -\mathcal{D}_S)\right]$$

so that $L_{Smth} > U_{Smth}$, indicating that the estimated bound is empty.

Now I discuss inference for $[L_{Smth}, U_{Smth}]$. Define the estimator of $[L_{Smth}, U_{Smth}]$ as

$$[\widehat{L}_{Smth}, \widehat{U}_{Smth}] = \left[\widehat{\mathcal{B}}_S - \frac{1}{2}\left(s(\widehat{\mathcal{E}}_S, \widehat{\mathcal{D}}_S) - s(\widehat{\mathcal{E}}_S, -\widehat{\mathcal{D}}_S)\right),\ \widehat{\mathcal{B}}_S + \frac{1}{2}\left(s(\widehat{\mathcal{E}}_S, \widehat{\mathcal{D}}_S) - s(\widehat{\mathcal{E}}_S, -\widehat{\mathcal{D}}_S)\right)\right].$$

Assumption 10 and the Delta method (Van der Vaart, 2000, Section 3) then imply that $(\widehat{L}_{Smth}, \widehat{U}_{Smth})$ is asymptotically normal:

$$\sqrt{N}((\widehat{L}_{Smth}, \widehat{U}_{Smth})' - (L_{Smth}, U_{Smth})') \xrightarrow{d} N\left(0, \begin{bmatrix} \sigma_L^2 & \rho\sigma_L\sigma_U \\ \rho\sigma_L\sigma_U & \sigma_U^2 \end{bmatrix}\right)$$

for some σ_L, σ_U, and ρ. This verifies Assumption 1 of Stoye (2020). Note that σ_L, σ_U, and ρ can be consistently estimated by bootstrap (Van der Vaart, 2000, Section 23).

Next, define the pseudo-true parameter (Stoye, 2020; Andrews and Kwon, 2024):

$$\mu_e^* = \frac{\sigma_L L_{Smth} + \sigma_U U_{Smth}}{\sigma_L + \sigma_U}.$$

Note that μ_e^* is well-defined even if $\mathcal{D}_S < 0$ and that $\mu_e^* \approx \mathcal{B}_S$. Define its estimator as

$$\widehat{\mu}_e^* = \frac{\widehat{\sigma}_L \widehat{L}_{Smth} + \widehat{\sigma}_U \widehat{U}_{Smth}}{\widehat{\sigma}_L + \widehat{\sigma}_U},$$

where $\widehat{\sigma}_L$, $\widehat{\sigma}_U$, and $\widehat{\rho}$ are consistent estimators of σ_L, σ_U, and ρ, respectively. Both μ_e^* and $\widehat{\mu}_e^*$ are well-defined under overidentification or misspecification.

Then, the $(1-\alpha)$-level confidence interval for μ_e is constructed as follows. First, consider an interval for μ_e based on the smoothed bounds $[\widehat{L}_{Smth}, \widehat{U}_{Smth}]$:

$$I_{\mu_e} = \left[\widehat{L}_{Smth} - \widehat{c}(\alpha)\frac{\widehat{\sigma}_L}{\sqrt{N}},\ \widehat{U}_{Smth} + \widehat{c}(\alpha)\frac{\widehat{\sigma}_U}{\sqrt{N}}\right],$$

where $\widehat{c}(\alpha)$ is the critical value specified in Table 1 of Stoye (2020). For instance, if $\alpha = 0.05$, then $\widehat{c}(0.05) = 1.64$ if $\widehat{\rho} < 0.8$, and $\widehat{c}(0.05) = 1.96$ if $\widehat{\rho} \approx 1$. Note that I_{μ_e} may be empty under overidentification or misspecification. Second, consider an interval for the pseudo-true parameter μ_e^*:

$$I_{\mu_e^*} = \left[\widehat{\mu}_e^* - \Phi\left(1 - \frac{\alpha}{2}\right)\frac{\widehat{\sigma}^*}{\sqrt{N}},\ \widehat{\mu}_e^* + \Phi\left(1 - \frac{\alpha}{2}\right)\frac{\widehat{\sigma}^*}{\sqrt{N}}\right]$$

where Φ is the standard normal CDF and

$$\widehat{\sigma}^* = \frac{\widehat{\sigma}_L \widehat{\sigma}_U \sqrt{2 + 2\widehat{\rho}}}{\widehat{\sigma}_L + \widehat{\sigma}_U}.$$

Then, the $(1-\alpha)$-level confidence interval for μ_e that is valid under overidentification and misspecification is given by

$$CI_{\mu_e} = I_{\mu_e} \cup I_{\mu_e^*}.$$

Theorem 1 of Stoye (2020) establishes the validity of CI_{μ_e}. Under overidentification, CI_{μ_e} asymptotically achieves the $(1-\alpha)$ coverage rate for the true parameter μ_e, where overidentification is resolved as $N \to \infty$. Under misspecification, CI_{μ_e} asymptotically achieves the coverage rate for the pseudo-true parameter μ_e^*.

5.2 Inference for the general parameters

I now discuss construction of a confidence interval for a general parameter θ in Section 4. By Theorem 2, the bounds $[L, U]$ of θ are given by

$$L = \max_{\lambda \in \mathbb{R}^K} \mathbb{E}(G_L(\lambda, W_i)), \quad \text{and} \quad U = \min_{\lambda \in \mathbb{R}^K} \mathbb{E}(G_U(\lambda, W_i)).$$

Note first that any $\theta \in [L, U]$ must satisfy

$$\begin{aligned} \theta \geq L &= \max_{\lambda \in \mathbb{R}^K} \mathbb{E}(G_L(\lambda, W_i)), \\ \theta \leq U &= \min_{\lambda \in \mathbb{R}^K} \mathbb{E}(G_U(\lambda, W_i)). \end{aligned}$$

For regularity of the inference procedure, consider a large compact set $R^K \subseteq \mathbb{R}^K$ and consider the inequalities

$$\begin{aligned} \theta \geq \tilde{L} &= \max_{\lambda \in R^K} \mathbb{E}(G_L(\lambda, W_i)), \\ \theta \leq \tilde{U} &= \min_{\lambda \in R^K} \mathbb{E}(G_U(\lambda, W_i)). \end{aligned}$$

I choose the set R^K to be large enough so that both $\lambda_0^L = \text{argmax}_\lambda \, \mathbb{E}(G_L(\lambda, W_i))$ and $\lambda_0^U = \text{argmin}_\lambda \, \mathbb{E}(G_U(\lambda, W_i))$ lie in the interior of R^K, in which case $[L, U] = [\tilde{L}, \tilde{U}]$. Otherwise, $[\tilde{L}, \tilde{U}]$ becomes an outer identified set of $[L, U]$. I then rewrite the above as

$$\begin{aligned} \theta &\geq \mathbb{E}(G_L(\lambda, W_i)) \quad \text{for all } \lambda \in R^K, \\ \theta &\leq \mathbb{E}(G_U(\lambda, W_i)) \quad \text{for all } \lambda \in R^K. \end{aligned}$$

Equivalently, these can be written as the following moment inequalities:

$$\begin{aligned} \mathbb{E}(G_L(\lambda, W_i) - \theta) &\leq 0 \quad \text{for all } \lambda \in R^K, \\ \mathbb{E}(\theta - G_U(\lambda, W_i)) &\leq 0 \quad \text{for all } \lambda \in R^K, \end{aligned} \tag{18}$$

which is a moment inequalities model with infinitely many restrictions indexed by λ.

The literature on many moment inequalities (Romano, Shaikh, and Wolf, 2014; Andrews and Shi, 2017; Chernozhukov, Chetverikov, and Kato, 2019; Bai, Santos, and Shaikh, 2022) develops procedures for constructing a confidence interval for θ. In this paper, I adopt the method of Andrews and Shi (2017), who provides an inference procedure for countably many moment inequalities. Note first that G_L is concave and G_U is convex in λ by Proposition 4, which implies that both functions are continuous in λ. This means that, for inference on θ, it suffices to consider:

$$\begin{aligned} \mathbb{E}(G_L(\lambda, W_i) - \theta) &\leq 0 \quad \text{for all } \lambda \in Q^K, \\ \mathbb{E}(\theta - G_U(\lambda, W_i)) &\leq 0 \quad \text{for all } \lambda \in Q^K, \end{aligned} \tag{19}$$

where $Q^K \subseteq R^K$ is a set of rational numbers in R^K. Section 9.2 of Andrews and Shi (2017) develops an inference procedure for this countable set of moment inequalities. The conditions for the validity of their procedure are given in Lemma 9.2 of Andrews and Shi (2017). Although this lemma is given for a single moment restriction with one-dimensional λ, its extension to two moment restrictions and to a K-dimensional λ is straightforward. Let $\mathbb{T} = Q^K \times \{0, 1\}$, and define a single moment function $h(W_i, \theta, \tau)$ for $\tau \in \mathbb{T}$ as

$$h(W_i, \theta, \tau) = \begin{cases} G_L(\lambda, W_i) - \theta & \text{if } \tau = (\lambda, 0), \\ \theta - G_U(\lambda, W_i) & \text{if } \tau = (\lambda, 1). \end{cases}$$

Then, since $\mathbb{T}$ is a countable set, its elements $\tau \in \mathbb{T}$ can be ordered as $\{\tau_j\}_{j=1}^{\infty}$, where j is one-dimensional. In what follows, I assume that the conditions of their Lemma 9.2 are satisfied, which are mild given Assumption 8 and the fact that Q^K is compact[4].

Assumption 11. There is $\underline{\sigma} > 0$ such that $\text{Var}(G_L(\lambda_0, W_i)) \geq \underline{\sigma}^2$ and $\text{Var}(G_U(\lambda_0, W_i)) \geq \underline{\sigma}^2$ for some fixed $\lambda_0 \in Q^K$. Also, there is a measurable function g such that $|G_L(\lambda_0, W_i)| \leq g(W_i)$, $|G_U(\lambda_0, W_i)| \leq g(W_i)$, and $\mathbb{E}((g(W_i)/\underline{\sigma})^{2+r}) \leq C$ for some $r > 0$ and $C < \infty$.

In what follows, I apply their inference method under assumptions of Theorem 2 and Assumption 11, where I choose the tuning parameters appropriately for brevity of dis-

[4] Lemma 9.2 of Andrews and Shi (2017) introduces a weight function associated with an ordering of the τ values in $\mathbb{T}$. This weight function does not affect the inference procedure, since it cancels out in the construction of the test statistics and therefore does not appear in any of the expressions. Consequently, for an ordering of the τ values, denoted by $\{\tau_j\}_{j=1}^{\infty}$, the weight function $w(j)$ can simply be set to be $w(j) = j^{-b}$ for $b > 0$ as suggested in Andrews and Shi (2017).

cussion. Given an i.i.d. sample $\{W_i\}_{i=1}^N$ of size N, define the sample quantities

$$\hat{\mu}_{G_L}(\lambda) = \frac{1}{N}\sum_{i=1}^N G_L(\lambda, W_i) \quad \text{and} \quad \hat{\sigma}_{G_L}(\lambda) = \sqrt{(1+\kappa)\frac{1}{N}\sum_{i=1}^N \left(G_L(\lambda, W_i) - \hat{\mu}_{G_L}(\lambda)\right)^2}$$

where $\kappa = 0.05$ is a small number, and where $\hat{\mu}_{G_U}(\lambda)$ and $\hat{\sigma}^2_{G_U}(\lambda)$ are defined similarly with G_U. Define the test statistic as

$$T_{AS}(\theta) = \sup_{\lambda \in Q^K} \max\left\{\frac{\sqrt{N}(\hat{\mu}_{G_L}(\lambda) - \theta)}{\hat{\sigma}_{G_L}(\lambda)}, \frac{\sqrt{N}(\theta - \hat{\mu}_{G_U}(\lambda))}{\hat{\sigma}_{G_U}(\lambda)}, 0\right\}^2,$$

which corresponds to the function S_3 in Andrews and Shi (2017). This test statistic is then compared to the critical value $c_{AS}(\alpha)$, which can be computed in two ways. First, to compute the plug-in asymptotic (PA) type critical value, define the two functions to be $\overline{\phi}_L(\theta, \lambda) = \overline{\phi}_U(\theta, \lambda) = 0$. In contrast, for the generalized moment selection (GMS) type critical value, define

$$\xi_L(\theta, \lambda_k) = \sqrt{\frac{N}{0.3 \ln N}} \times \frac{\hat{\mu}_{G_L}(\lambda) - \theta}{\hat{\sigma}_{G_L}(\lambda)},$$
$$\xi_U(\theta, \lambda_k) = \sqrt{\frac{N}{0.3 \ln N}} \times \frac{\theta - \hat{\mu}_{G_U}(\lambda)}{\hat{\sigma}_{G_U}(\lambda)},$$

and define the two functions

$$\overline{\phi}_L(\theta, \lambda) = \sqrt{\frac{0.4 \ln N}{\ln \ln N}} \times \hat{\sigma}_{G_L}(\lambda) \times \mathbf{1}(\xi_L(\theta, \lambda) < -1),$$
$$\overline{\phi}_U(\theta, \lambda) = \sqrt{\frac{0.4 \ln N}{\ln \ln N}} \times \hat{\sigma}_{G_U}(\lambda) \times \mathbf{1}(\xi_U(\theta, \lambda) < -1).$$

Let $\{W_i^{(b)}\}_{i=1}^N$ be the empirical bootstrap sample of $\{W_i\}_{i=1}^N$, meaning that each $\{W_i^{(b)}\}_{i=1}^N$ is drawn from $\{W_i\}_{i=1}^N$ with replacement. Let $\hat{\mu}^{(b)}_{G_L}$, $\hat{\mu}^{(b)}_{G_U}$, $\hat{\sigma}^{(b)}_{G_L}$, $\hat{\sigma}^{(b)}_{G_U}$ be the values of $\hat{\mu}_{G_L}$, $\hat{\mu}_{G_U}$, $\hat{\sigma}_{G_L}$, $\hat{\sigma}_{G_U}$ computed with $\{W_i^{(b)}\}_{i=1}^N$ instead of $\{W_i\}_{i=1}^N$. Then, compute the statistic

$$c^{(b)}_{AS}(\theta) = \sup_{\lambda \in Q^K} \max\left\{\frac{\sqrt{N}(\hat{\mu}^{(b)}_{G_L}(\lambda) - \hat{\mu}_{G_L}(\lambda)) - \overline{\phi}_L(\theta, \lambda)}{\hat{\sigma}^{(b)}_{G_L}(\lambda)}, \right.$$
$$\left. \frac{\sqrt{N}(\hat{\mu}_{G_U}(\lambda) - \hat{\mu}^{(b)}_{G_U}(\lambda)) - \overline{\phi}_U(\theta, \lambda)}{\hat{\sigma}^{(b)}_{G_U}(\lambda)}, 0\right\}^2.$$

The critical value $c_{AS}(\theta, \alpha)$ is then defined as the $(1-\alpha)$ quantile of the bootstrapped $c^{(b)}_{AS}$

values. The confidence set for θ is then given by $\{\theta \mid T_{AS}(\theta) \leq c_{AS}(\theta, \alpha)\}$. Note that, for the PA type, the critical value $c_{AS}(\theta, \alpha)$ does not depend on θ since $\overline{\phi}_L(\theta, \lambda) = \overline{\phi}_U(\theta, \lambda) = 0$. Consequently, the PA type confidence set simplifies to the interval

$$\left[\sup_{\lambda} \left\{\hat{\mu}_{G_L}(\lambda) - \sqrt{c_{AS}(\alpha)} \times \frac{\hat{\sigma}_{G_L}(\lambda)}{\sqrt{N}}\right\}, \ \inf_{\lambda} \left\{\hat{\mu}_{G_U}(\lambda) + \sqrt{c_{AS}(\alpha)} \times \frac{\hat{\sigma}_{G_U}(\lambda)}{\sqrt{N}}\right\}\right].$$

For the GMS type, $c_{AS}(\theta, \alpha)$ depends on θ only through $\overline{\phi}_L(\theta, \lambda)$ and $\overline{\phi}_U(\theta, \lambda)$, so one only needs to update these two quantities when searching for θ that satisfy $T_{AS}(\theta) \leq c_{AS}(\theta, \alpha)$.

When K is large, searching for supremum over all $\lambda \in Q^K$ in $T_{AS}(\theta)$ and $c_{AS}^{(b)}(\theta)$ can be computationally prohibitive. However, note that the inequalities in (19) bind only at two λ values, namely at $\lambda_L^* = \operatorname{argmax}_{\lambda} \mathbb{E}(G_L(\lambda, W_i))$ and $\lambda_U^* = \operatorname{argmin}_{\lambda} \mathbb{E}(G_U(\lambda, W_i))$. Moreover, since G_L is concave and G_U is convex, the inequalities become loose for λ values that are far from λ_L^* and λ_U^*. Consequently, in practice, one can focus the search for λ on neighborhoods of λ_L^* and λ_U^*. This intuition is formalized in the GMS type critical value, where $\overline{\phi}_L(\theta, \lambda)$ and $\overline{\phi}_U(\theta, \lambda)$ become positive when

$$\hat{\mu}_{G_L}(\lambda) - \sqrt{\frac{N}{0.3 \ln N}} \times \frac{\hat{\sigma}_{G_L}(\lambda)}{\sqrt{N}} < \theta, \quad \text{and} \quad \theta < \hat{\mu}_{G_U}(\lambda) + \sqrt{\frac{N}{0.3 \ln N}} \times \frac{\hat{\sigma}_{G_U}(\lambda)}{\sqrt{N}},$$

respectively. This means that the λ values for which $\hat{\mu}_{G_L}(\lambda)$ is too small or $\hat{\mu}_{G_U}(\lambda)$ is too large are not likely to attain the supremum.

While λ_L^* and λ_U^* are population quantities, they can be approximated with their sample analogues. Consider the estimators of L and U in Theorem 2 obtained by replacing expectations with the sample means:

$$\hat{L}_N = \max_{\lambda \in \mathbb{R}^K} \frac{1}{N} \sum_{i=1}^{N} G_L(\lambda, W_i), \tag{20}$$

and

$$\hat{U}_N = \min_{\lambda \in \mathbb{R}^K} \frac{1}{N} \sum_{i=1}^{N} G_U(\lambda, W_i). \tag{21}$$

Define the λ values that attain the optima as

$$\hat{\lambda}_N^L = \operatorname*{argmax}_{\lambda \in \mathbb{R}^K} \frac{1}{N} \sum_{i=1}^{N} G_L(\lambda, W_i), \quad \text{and} \quad \hat{\lambda}_N^U = \operatorname*{argmax}_{\lambda \in \mathbb{R}^K} \frac{1}{N} \sum_{i=1}^{N} G_U(\lambda, W_i).$$

The search for supremum can then be focused on the neighborhoods of $\hat{\lambda}_N^L$ and $\hat{\lambda}_N^U$.

The procedure naturally extends to a vector-valued parameter $\theta \in \mathbb{R}^d$, by considering (18) on each component of θ. For example, the moment inequalities for $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ are:

$$\begin{aligned}
\mathbb{E}(G_{L1}(\lambda, W_i) - \theta_1) &\leq 0 \quad \text{for all } \lambda \in Q^K, \\
\mathbb{E}(\theta_1 - G_{U1}(\lambda, W_i)) &\leq 0 \quad \text{for all } \lambda \in Q^K, \\
\mathbb{E}(G_{L2}(\lambda, W_i) - \theta_2) &\leq 0 \quad \text{for all } \lambda \in Q^K, \\
\mathbb{E}(\theta_2 - G_{U2}(\lambda, W_i)) &\leq 0 \quad \text{for all } \lambda \in Q^K,
\end{aligned} \tag{22}$$

where G_{Uk} and G_{Lk} denote the functions G_L and G_U in (18) corresponding to θ_k for $k = 1, 2$. Applying the same inference procedure then yields a confidence region in $\mathbb{R}^2$. This extension can be used to construct a confidence interval for the variance of random coefficients, which involves both first and second moments. Alternatively, it can be constructed by the Bonferroni correction to the individual bounds.

Lastly, I discuss overidentification and model misspecification in inference for the general parameters. Under overidentification or misspecification, the finite-sample optimizers $\hat{\lambda}_N^L$ and $\hat{\lambda}_N^U$, as well as the test statistic $T_{AS}(\theta)$ and its bootstrap critical value $c_{AS}^{(b)}(\theta)$, all diverge to $+\infty$. In contrast to the case of mean parameters, it is substantially more challenging to deal with these issues for general parameters. Andrews and Kwon (2024) develop a general method for constructing valid confidence intervals under overidentification or misspecification, but their approach applies to a finite number of moment restrictions and therefore is not readily applicable to the countably infinite set considered here. Extending their approach to countably many moment restrictions is beyond the scope of this paper and is not pursued here. Instead, in Online Appendix B.5, I discuss a heuristic modification of the procedure of Andrews and Shi (2017), where I introduce alternative finite-sample optimizers $\tilde{\lambda}_N^L$ and $\tilde{\lambda}_N^U$ that remain well-defined and finite under overidentification, which closely align with the spirit of Andrews and Kwon (2024). Using these optimizers, I implement the procedure of Andrews and Shi (2017) by calculating the supremum in the neighborhoods of $\tilde{\lambda}_N^L$ and $\tilde{\lambda}_N^U$. I check the performance of this heuristic procedure via simulation, also in Online Appendix B.5.

6 Application to lifecycle earnings dynamics

6.1 Overview

Lifecycle earnings dynamics serve as a key input in various macroeconomic models. For example, in models of consumption and savings dynamics (Hall and Mishkin, 1982; Blundell, Pistaferri, and Preston, 2008; Blundell, Pistaferri, and Saporta-Eksten, 2016; Arellano, Blundell, and Bonhomme, 2017), households facing a higher earnings risk accumulate more precautionary savings to smooth consumption over time. As Guvenen (2009) points out, specifying an earnings process that highlights features of real data is important for properly calibrating and drawing conclusions from these models.

When used as an input, it is common to specify earnings dynamics using a parsimonious linear model. Guvenen (2007, 2009) studied two leading views on parsimonious specification of the earnings dynamics. Consider two earnings processes[5]:

$$\begin{aligned} Y_{it} &= \alpha_i + z_{it}, & z_{it} &= \rho z_{i,t-1} + \eta_{it}, & &\text{(RIP)} \\ Y_{it} &= \alpha_i + \beta_i h_{it} + z_{it}, & z_{it} &= \rho z_{i,t-1} + \eta_{it}, & &\text{(HIP)} \end{aligned} \tag{23}$$

where $h = \text{age} - \max\{\text{years of schooling}, 12\} - 6$ is potential years of experience, Y_{it} is the residual log-earnings obtained by regressing log-earnings on time indicators and their interactions with a cubic polynomial in h, and (α_i, β_i) are heterogeneous coefficients. In addition, $\{z_{it}\}$ is an AR(1) process with a mean zero shock η_{it}[6]. These two models are known as the Restricted Income Profiles (RIP) process and the Heterogeneous Income Profiles (HIP) process, respectively. In both models, ρ captures the earnings persistence that households face. As Guvenen (2009) summarizes, the literature reports $0.5 < \rho < 0.7$ and $\text{Var}(\beta_i) > 0$ for the HIP process (e.g., Lillard and Weiss, 1979; Baker, 1997), meaning that households experience modest persistence and heterogeneous trends. By contrast, MaCurdy (1982) tested the hypothesis that $\text{Var}(\beta_i) = 0$ and did not reject it. The literature reports $\rho \approx 1$ for the RIP process (e.g., Abowd and Card, 1989; Topel and Ward, 1992), meaning households experience extreme persistence and homogeneous trends. Guvenen (2007) demonstrated that the HIP process better aligns with features of consumption data, and Guvenen (2009) showed that misspecifying the HIP process as a RIP process leads to an upward biased estimate of ρ, often obtaining $\rho \approx 1$.

While there is vast literature on unobserved heterogeneity in β_i and its influence on ρ, relatively few studies examines heterogeneity in ρ itself. Notable recent studies include Browning, Ejrnaes, and Alvarez (2010), Alan, Browning, and Ejrnæs (2018), and Pesaran and Yang (2024); the first two assume a factor structure for ρ_i, and the latter imposes stationarity of (23) and assumes η_{it} are i.i.d. over i and t. In this section, I estimate a

[5] As Guvenen (2007) points out, this is a stylized version of what is used in the literature, but it still captures features important for the discussion.

[6] In the literature, it is standard to add a transitory income process to (23). I present estimation results that account for a transitory income process in Online Appendix B.6. The estimation results yield similar qualitative conclusions outlined in this subsection.

generalization of (23) where ρ varies across individuals, writing $\rho = \rho_i$, where the distribution of ρ_i and its correlation with $(\alpha_i, \beta_i, Y_{i0})$ are unrestricted. Differently from Pesaran and Yang (2024), who also extend Guvenen (2009), the distribution of η_{it} also remains unrestricted and may depend on $(\alpha_i, \beta_i, \rho_i)$, allowing for heteroskedasticity.

In the remainder of this section, I find that, when ρ is allowed to vary across individuals, both RIP and HIP specifications deliver similar estimates of $\mathbb{E}(\rho_i)$ that are significantly less than 1. At the 95% confidence level, the upper bounds of the confidence intervals for $\mathbb{E}(\rho_i)$ under both processes are between 0.65 and 0.70, and the two intervals have substantial overlap. This result suggests that, when ρ is allowed to be heterogeneous, choosing RIP over HIP or vice versa may not lead to serious misspecification of ρ_i. Moreover, the 90% confidence intervals for $\mathrm{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$ for $r \in (0, 1)$ in the RIP model suggest substantial heterogeneity in ρ_i. In particular, the lower confidence limit for $\mathrm{Var}(\rho_i)$ is 0.052, implying a standard deviation of 0.228, and the confidence intervals for the CDF of ρ_i suggest that at least 40% of individuals have $\rho_i \leq 0.8$.

6.2 Data and models

I analyze data on U.S. households from the Panel Study of Income Dynamics (PSID) dataset. I use the dataset of Guvenen (2009), who analyzed the PSID dataset of male heads of households collected annually. The dataset consists of male head of households who are not in the poverty (SEO) subsample and who consecutively reported positive hours (between 520 and 5110 hours a year) and earnings (between a preset minimum and maximum wage). From the dataset of Guvenen (2009), I select individuals observed consecutively from 1976 to 1991, yielding $N = 800$ and $T = 15$, where the first wave serves as the initial value of earnings. I estimate two dynamic random coefficient models:

$$\begin{aligned} Y_{it} &= \alpha_i + \rho_i Y_{i,t-1} + \eta_{it}, & \mathbb{E}(\eta_{it} | \alpha_i, \rho_i, Y_i^{t-1}) &= 0, & \text{(RIP-RC)} \\ Y_{it} &= \alpha_i + \beta_i h_{it} + \rho_i Y_{i,t-1} + \eta_{it}, & \mathbb{E}(\eta_{it} | \alpha_i, \beta_i, \rho_i, Y_i^{t-1}, h_i) &= 0. & \text{(HIP-RC)} \end{aligned} \qquad (24)$$

These models generalize (23), and they can be derived by quasi-differencing Y_{it} in (23) and assuming $h_{it} \approx h_{i,t-1} + 1$. Specifically, quasi-differencing the RIP process gives

$$Y_{it} = \alpha_i(1 - \rho_i) + \rho_i Y_{i,t-1} + \eta_{it} \equiv \tilde{\alpha}_i + \rho_i Y_{i,t-1} + \eta_{it}.$$

Likewise, quasi-differencing the HIP process gives

$$Y_{it} = \alpha_i(1 - \rho_i) + \beta_i \rho_i + \beta_i(1 - \rho_i) h_{it} + \rho_i Y_{i,t-1} + \eta_{it} \equiv \tilde{\alpha}_i + \tilde{\beta}_i h_{it} + \rho_i \tilde{Y}_{i,t-1} + \eta_{it}.$$

Note that Guvenen (2009) defines Y_{it} in (23) as the residual from the regression on time indicators and their interactions with a cubic polynomial in h_{it}, i.e., the regression

$$\begin{aligned} Y_{it} &= \sum_{s=1976}^{1991} \Big(\mathbf{1}(t=s)\delta_{0,s} + \mathbf{1}(t=s)h_{it}\delta_{1,s} + \mathbf{1}(t=s)h_{it}^2\delta_{2,s} + \mathbf{1}(t=s)h_{it}^3\delta_{3,s} \Big) + \nu_{it} \\ &\equiv X_{it}'\delta + \nu_{it}, \end{aligned} \tag{25}$$

where Y_{it} is now the raw log-earnings data, and X_{it} and δ denote the regressors and the coefficients in (25), i.e., $X_{it} = \text{vec}\left\{ (\mathbf{1}(t=s), \mathbf{1}(t=s)h_{it}, \mathbf{1}(t=s)h_{it}^2, \mathbf{1}(t=s)h_{it}^3)_{s=1976}^{1991} \right\}$ and $\delta = \text{vec}\left\{ (\delta_{0,s}, \delta_{1,s}, \delta_{2,s}, \delta_{3,s})_{s=1976}^{1991} \right\}$. Guvenen (2009) estimate the RIP and HIP models in (23) using the two-step procedure that is standard in the literature, where one first obtains the residuals from the regression in (25), and then one treats these residuals as Y_{it} and estimate the RIP and HIP models in (23). The motivation of this approach is to first "partial out" the control variables X_{it} and then consider earnings dynamics that are free of X_{it}. However, this approach may understate the standard errors of the RIP and HIP estimates, since it fails to account for the sampling variability introduced by the first-stage regression. Moreover, any estimation error in the first stage may appear as heterogeneity in (α_i, β_i) in the RIP-RC and HIP-RC specifications. To address these issues, I also consider a joint model of the control variables term in (25) and the RIP-RC and HIP-RC models in (24). Specifically, I estimate

$$\begin{aligned} Y_{it} &= X_{it}'\delta + \alpha_i + \rho_i Y_{i,t-1} + \eta_{it}, & \mathbb{E}(\eta_{it}|\alpha_i, \rho_i, Y_i^{t-1}, X_i) &= 0, & \text{(RIP-RC-J)} \\ Y_{it} &= X_{it}'\delta + \alpha_i + \beta_i h_{it} + \rho_i Y_{i,t-1} + \eta_{it}, & \mathbb{E}(\eta_{it}|\alpha_i, \beta_i, \rho_i, Y_i^{t-1}, h_i, X_i) &= 0. & \text{(HIP-RC-J)} \end{aligned} \tag{26}$$

where the homogeneous coefficients δ and the heterogeneous coefficients $(\alpha_i, \beta_i, \rho_i)$ are jointly considered. I refer to these specifications in (26) as RIP-RC-J and HIP-RC-J. I estimate the mean parameters of (26) using the bounds in Proposition 3.

Note that, for estimation of the RIP-RC-J model, the regressor $\mathbf{1}(t=1976)$ must be removed from the model because it is multicollinear with the individual-specific intercept α_i. Likewise, for estimation of the HIP-RC-J model, both $\mathbf{1}(t=1976)$ and $\mathbf{1}(t=1976)h_{it}$ must be dropped from X_{it} to avoid multicollinearity with the α_i and the h_{it} terms. These exclusions ensure that the no-multicollinearity condition of Assumption 5 holds. After these removals, X_{it} has 59 regressors in RIP-RC-J and 58 in HIP-RC-J models. The estimation result below will show that the bounds in Proposition 3 produces informative confidence intervals under this setup, demonstrating its practical applicability with a large number of regressors with homogeneous coefficients.

In what follows, I construct confidence intervals for $\mathbb{E}(\rho_i)$ under the RIP-RC, HIP-

RC, RIP-RC-J, and HIP-RC-J specifications, using the bounds presented in Propositions 2 and 3. For the RIP-RC and HIP-RC models, I employ the two-step procedure of Guvenen (2009) which first obtains residuals from (25) and then uses these residuals as Y_{it} in the RIP-RC and HIP-RC models. In contrast, for the RIP-RC-J and HIP-RC-J models, I directly estimate $\mathbb{E}(\rho_i)$ from (26), jointly considering the control variables term. In addition, I construct the confidence intervals for $\text{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$ over the grid $r \in \{0.1, \ldots, 0.9\}$ under the RIP-RC model, using the bounds presented in Section 4.2 and assuming $\mathcal{B} = [-3, 3] \times [0, 1]$ as the support of (α_i, ρ_i).

For calculation of the bounds, I choose $S_{it} = (1, Y_{i,\max\{0,\, t-5\}}, \ldots, Y_{i,t-1})'$ for RIP-type models, and $S_{it} = (1, Y_{i,\max\{0,\, t-5\}}, \ldots, Y_{i,t-1}, h_{i,\max\{1,\, t-5\}}, \ldots, h_{i,\min\{T,\, t+5\}})$ for HIP-type models. For these choices of S_{it}, the overidentification issue arises in the estimated bounds. For inference on $\mathbb{E}(\rho_i)$, I apply the procedure of Stoye (2020) discussed in Section 5.1. For inference on $\text{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$, I adopt the heuristic modification of Andrews and Shi (2017) described in Online Appendix B.5. Guided by simulation results, I evaluate the supremum with 100 grid points in the neighborhoods. All critical values are calculated with 1000 bootstrap replications, using the PA type for $\text{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$. The interval for $\text{Var}(\rho_i)$ is constructed with the Bonferroni correction.

6.3 Results

The 95% confidence intervals for $\mathbb{E}(\rho_i)$ are reported in Table 2. Both models estimate $\mathbb{E}(\rho_i)$ to be significantly less than 1, and the confidence intervals in RIP-RC and HIP-RC demonstrate substantial overlap, having similar upper confidence limits. This suggests that specifying homogeneous or heterogeneous β_i does not lead to serious misspecification when ρ_i is allowed to be heterogeneous. In addition, the estimates from RIP-RC-J and HIP-RC-J are slightly higher than their counterparts of RIP-RC and HIP-RC, suggesting that the estimation error from the first stage may appear as an additional term within ρ_i that attenuates the magnitude of ρ_i. Note that these intervals are computed with the procedure described in Section 5.1, which is robust to overidentification and model misspecification. These findings are qualitatively similar when also considering the transitory income process, as reported in Online Appendix B.6.

The 90% confidence intervals for $\text{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$ over the grid $r \in \{0.1, \ldots, 0.9\}$ for the RIP-RC model are reported in Table 3. The lower confidence limit of $\text{Var}(\rho_i)$ is 0.052, implying a standard deviation of 0.228, suggesting substantial heterogeneity in ρ_i. Similar evidence is observed from confidence intervals for the CDF of ρ_i. They indicate that at least 40% of households have $\rho_i \leq 0.8$ and at least 28% have $\rho_i \leq 0.5$. These find-

Parameter	RIP-RC	HIP-RC	RIP-RC-J	HIP-RC-J
$\mathbb{E}(\rho_i)$	[0.485, 0.651]	[0.346, 0.653]	[0.545, 0.674]	[0.375, 0.659]

Table 2: Confidence intervals of $\mathbb{E}(\rho_i)$ for the RIP type and the HIP type processes with heterogeneous coefficients. The nominal coverage probability is 0.95. These confidence intervals are robust to overidentification and model misspecification.

Parameter	RIP-RC
$\text{Var}(\rho_i)$	[0.052, 1.257]
$\mathbb{P}(\rho_i \leq 0.1)$	[0.074, 0.989]
$\mathbb{P}(\rho_i \leq 0.2)$	[0.120, 0.992]
$\mathbb{P}(\rho_i \leq 0.3)$	[0.186, 1.000]
$\mathbb{P}(\rho_i \leq 0.4)$	[0.193, 1.000]
$\mathbb{P}(\rho_i \leq 0.5)$	[0.286, 1.000]
$\mathbb{P}(\rho_i \leq 0.6)$	[0.324, 1.000]
$\mathbb{P}(\rho_i \leq 0.7)$	[0.383, 1.000]
$\mathbb{P}(\rho_i \leq 0.8)$	[0.409, 1.000]
$\mathbb{P}(\rho_i \leq 0.9)$	[0.439, 1.000]

Table 3: Confidence intervals of $\text{Var}(\rho_i)$ and $\mathbb{P}(\rho_i \leq r)$ for the RIP and the HIP processes with heterogeneous coefficients. The nominal coverage probability is 0.90.

ings suggest substantial unobserved heterogeneity in the earnings risk that households face, highlighting the importance of allowing for heterogeneity in ρ_i in modeling income processes that reflect features of real data.

7 Conclusion

This paper studies the identification and estimation of dynamic random coefficient models in a short panel context. The model extends the standard dynamic panel linear model with fixed effects (Arellano and Bond, 1991; Blundell and Bond, 1998), allowing coefficients to be individual-specific. I show that the model is not point-identified but rather partially identified, and I characterize the identified sets of the mean, variance and CDF of the random coefficients using the dual representation of an infinite-dimensional linear program. I propose a computationally tractable estimation and inference procedure by adopting the approach of Stoye (2020) for the mean parameters and Andrews and Shi (2017) for the variance and CDF parameters. The procedure of Stoye (2020) is robust to overidentification and model misspecification.

I use my method to estimate unobserved heterogeneity in earnings persistence across U.S. households using the PSID dataset. I find that the average earnings persistence is significantly less than 1 when it is allowed to be heterogeneous. Moreover, its confidence interval under the RIP and HIP specifications show substantial overlap, suggesting that choosing RIP over HIP or vice versa does not lead to serious misspecification about the earnings process when persistence is heterogeneous. Lastly, confidence intervals for the variance and CDF of the earnings persistence parameter suggest a substantial degree of unobserved heterogeneity, which is a key source of heterogeneity in consumption and savings behaviors.

8 Acknowledgements

This is based on my PhD dissertation at the University of Chicago. I am deeply indebted to Stéphane Bonhomme, Alexander Torgovitsky, and Guillaume Pouliot for their invaluable guidance and support. I thank Manuel Arellano, Timothy Armstrong, Antonio Galvao, Greg Kaplan, Roger Koenker, Zhipeng Liao, Jack Light, Hashem Pesaran, Azeem Shaikh, Shuyang Sheng, Panagiotis Toulis, and Ying Zhu for helpful discussions and comments. I also thank participants of the Econometrics Working Group at the University of Chicago.

References

Abowd, John M and David Card. 1989. "On the covariance structure of earnings and hours changes." *Econometrica* 57 (2):411–445.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica* 83 (6):2411–2451.

Alan, Sule, Martin Browning, and Mette Ejrnæs. 2018. "Income and consumption: A micro semistructural analysis with pervasive heterogeneity." *Journal of Political Economy* 126 (5):1827–1864.

Anderson, Edward J. 1983. "A review of duality theory for linear programming over topological vector spaces." *Journal of Mathematical Analysis and Applications* 97 (2):380–392.

Andrews, Donald WK and Soonwoo Kwon. 2024. "Misspecified moment inequality models: Inference and diagnostics." *Review of Economic Studies* 91 (1):45–76.

Andrews, Donald WK and Xiaoxia Shi. 2013. "Inference based on conditional moment inequalities." *Econometrica* 81 (2):609–666.

———. 2017. "Inference based on many conditional moment inequalities." *Journal of Econometrics* 196 (2):275–287.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and consumption dynamics: a nonlinear panel data framework." *Econometrica* 85 (3):693–734.

Arellano, Manuel and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies* 58 (2):277–297.

Arellano, Manuel and Stéphane Bonhomme. 2012. "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies* 79 (3):987–1020.

———. 2021. "Recovering latent variables by matching." *Journal of the American Statistical Association* :1–14.

Bai, Yuehao, Andres Santos, and Azeem M Shaikh. 2022. "A two-step method for testing many moment inequalities." *Journal of Business & Economic Statistics* 40 (3):1070–1080.

Baker, Michael. 1997. "Growth-rate heterogeneity and the covariance structure of life-cycle earnings." *Journal of Labor Economics* 15 (2):338–375.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari. 2011. "Sharp identification regions in models with convex moment predictions." *Econometrica* 79 (6):1785–1821.

Bierens, Herman J. 1990. "A consistent conditional moment test of functional form." *Econometrica* 58 (6):1443–1458.

Blundell, Richard and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87 (1):115–143.

Blundell, Richard, Hamish Low, and Ian Preston. 2013. "Decomposing changes in income risk using consumption data." *Quantitative Economics* 4 (1):1–37.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption inequality and partial insurance." *American Economic Review* 98 (5):1887–1921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption inequality and family labor supply." *American Economic Review* 106 (2):387–435.

Browning, Martin, Mette Ejrnaes, and Javier Alvarez. 2010. "Modelling income processes with lots of heterogeneity." *Review of Economic Studies* 77 (4):1353–1381.

Chamberlain, Gary. 1992. "Efficiency bounds for semiparametric regression." *Econometrica* 60 (3):567–596.

———. 1993. "Feedback in panel data models." *Working paper* .

———. 2022. "Feedback in panel data models." *Journal of Econometrics* 226 (1):4–20.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2019. "Inference on causal and structural parameters using many moment inequalities." *Review of Economic Studies* 86 (5):1867–1900.

Galichon, Alfred and Marc Henry. 2009. "A test of non-identifying restrictions and confidence regions for partially identified parameters." *Journal of Econometrics* 152 (2):186–196.

Graham, Bryan S and James L Powell. 2012. "Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models." *Econometrica* 80 (5):2105–2152.

Gu, Jiaying and Roger Koenker. 2017. "Unobserved heterogeneity in income dynamics: An empirical Bayes perspective." *Journal of Business & Economic Statistics* 35 (1):1–16.

Gunsilius, Florian. 2019. "Bounds in continuous instrumental variable models." *Working paper* .

Guvenen, Fatih. 2007. "Learning your earning: Are labor income shocks really very persistent?" *American Economic Review* 97 (3):687–712.

———. 2009. "An empirical investigation of labor income processes." *Review of Economic dynamics* 12 (1):58–79.

Hall, Robert E and Frederic S Mishkin. 1982. "The sensitivity of consumption to transitory income: Estimates from panel data on households." *Econometrica* 50 (2):461–481.

Honoré, Bo E and Adriana Lleras-Muney. 2006. "Bounds in competing risks models and the war on cancer." *Econometrica* 74 (6):1675–1698.

Honoré, Bo E and Elie Tamer. 2006. "Bounds on parameters in panel dynamic discrete choice models." *Econometrica* 74 (3):611–629.

Kaplan, Greg and Giovanni L Violante. 2014. "A model of the consumption response to fiscal stimulus payments." *Econometrica* 82 (4):1199–1239.

Kasahara, Hiroyuki, Paul Schrimpf, and Michio Suzuki. 2023. "Identification and estimation of production function with unobserved heterogeneity." *arXiv preprint arXiv:2305.12067* .

Kiefer, Jack. 1959. "Optimum experimental designs." *Journal of the Royal Statistical Society: Series B* 21 (2):272–304.

Lasserre, Jean-Bernard. 2010. *Moments, positive polynomials and their applications*. World Scientific.

———. 2015. *An introduction to polynomial and semi-algebraic optimization*. Cambridge University Press.

Levinsohn, James and Amil Petrin. 2003. "Estimating production functions using inputs to control for unobservables." *Review of Economic Studies* 70 (2):317–341.

Li, Lixiong. 2018. "Identification of structural and counterfactual parameters in a large class of structural econometric models." *Working paper* .

Lillard, Lee A and Yoram Weiss. 1979. "Components of variation in panel earnings data: American scientists 1960-70." *Econometrica* 47 (2):437–454.

MaCurdy, Thomas E. 1982. "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." *Journal of Econometrics* 18 (1):83–114.

Meghir, Costas and Luigi Pistaferri. 2004. "Income variance dynamics and heterogeneity." *Econometrica* 72 (1):1–32.

Milgrom, Paul and Ilya Segal. 2002. "Envelope theorems for arbitrary choice sets." *Econometrica* 70 (2):583–601.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using instrumental variables for inference about policy relevant treatment parameters." *Econometrica* 86 (5):1589–1619.

Molinari, Francesca. 2008. "Partial identification of probability distributions with misclassified data." *Journal of Econometrics* 144 (1):81–117.

Olley, G Steven and Ariel Pakes. 1996. "The dynamics of productivity in the telecommunications equipment industry." *Econometrica* 64 (6):1263–1297.

Pesaran, M Hashem and Liying Yang. 2024. "Heterogeneous autoregressions in short T panel data models." *Journal of Applied Econometrics* 39 (7):1359–1378.

Romano, Joseph P, Azeem M Shaikh, and Michael Wolf. 2014. "A practical two-step method for testing moment inequalities." *Econometrica* 82 (5):1979–2002.

Schennach, Susanne M. 2014. "Entropic latent variable integration via simulation." *Econometrica* 82 (1):345–385.

Stoye, Jörg. 2020. "A simple, short, but never-empty confidence interval for partially identified parameters." *arXiv preprint arXiv:2010.10484* .

Topel, Robert H and Michael P Ward. 1992. "Job mobility and the careers of young men." *Quarterly Journal of Economics* 107 (2):439–479.

Torgovitsky, Alexander. 2019. "Nonparametric inference on state dependence in unemployment." *Econometrica* 87 (5):1475–1505.

Van der Vaart, Aad W. 2000. *Asymptotic statistics*. Cambridge university press.

Wooldridge, Jeffrey M. 2005. "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models." *Review of Economics and Statistics* 87 (2):385–390.

Appendices

A Online Appendix: Proofs

Proof of Proposition 1. To prove Proposition 1, I start from a data generating process that satisfies (7) and then modify it to construct a new process that is observationally equivalent.

Let $(\gamma_i, \beta_i, Y_{i0}, Y_{i1}, Y_{i2})$ be the random variables that satisfy (7). These variables must satisfy two conditions. First, they generate the observed data:

$$\mathbb{P}(Y_{i0} \leq y_0, Y_{i1} \leq y_1, Y_{i2} \leq y_2) = \mathbb{F}(y_0, y_1, y_2), \tag{27}$$

where $\mathbb{F}$ is the observed cumulative distribution function of (Y_{i0}, Y_{i1}, Y_{i2}). Second, they satisfy the model constraints:

$$\begin{aligned} \mathbb{E}(Y_{i1} - \gamma_i - \beta_i Y_{i0} | \gamma_i, \beta_i, Y_{i0}) &= 0 \qquad \text{for all } (\gamma_i, \beta_i, Y_{i0}), \\ \mathbb{E}(Y_{i2} - \gamma_i - \beta_i Y_{i1} | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) &= 0 \qquad \text{for all } (\gamma_i, \beta_i, Y_{i0}, Y_{i1}). \end{aligned} \tag{28}$$

Now, I construct a new data generating process that also satisfies (27) and (28). Given the joint distribution of $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$, let k be a deterministic function of $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$ such that

$$\mathbb{E}((Y_{i1} - Y_{i0})k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \gamma_i, \beta_i, Y_{i0}) = 0, \tag{29}$$

so that k is orthogonal to $Y_{i1} - Y_{i0}$ conditional on $(\gamma_i, \beta_i, Y_{i0})$. I will specify the explicit expression for k later in the proof. Then, given the function k, I define new random coefficients $(\tilde{\gamma}_i, \tilde{\beta}_i)$ as the following deterministic functions of $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$:

$$\left(\begin{array}{c} \tilde{\gamma}_i \\ \tilde{\beta}_i \end{array} \right) \Bigg| (Y_{i0} = y_0, Y_{i1} = y_1, \gamma_i = r, \beta_i = b) = \left(\begin{array}{c} r - y_1 k(r, b, y_0, y_1) \\ b + k(r, b, y_0, y_1) \end{array} \right). \tag{30}$$

I then define $\tilde{Y}_{i2}$ conditional on $(Y_{i0}, Y_{i1}, \gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i)$ as

$$\begin{aligned} &\tilde{Y}_{i2} | (Y_{i0} = y_0,\ Y_{i1} = y_1,\ \gamma_i = r,\ \beta_i = b,\ \tilde{\gamma}_i = \tilde{r},\ \tilde{\beta}_i = \tilde{b}) \\ &\overset{d}{=} Y_{i2} | (Y_{i0} = y_0,\ Y_{i1} = y_1,\ \gamma_i = r,\ \beta_i = b). \end{aligned} \tag{31}$$

Equations (30) and (31) together yield a joint distribution of $(\gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}, \tilde{Y}_{i2})$, which in turn implies a joint distribution of $(\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}, \tilde{Y}_{i2})$. I now show that this new data generating process, given by $(\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}, \tilde{Y}_{i2})$, satisfies both (27) and (28).

I first show that the new data generating process satisfies (27). Since the joint distribution of (Y_{i0}, Y_{i1}) remains unchanged, it suffices to verify that, conditional on $(Y_{i0} = y_0, Y_{i1} = y_1)$, the distribution of $\tilde{Y}_{i2}$ coincides with that of Y_{i2}. By the definition of $\tilde{Y}_{i2}$ in (31) and the law of iterated expectations, I obtain

$$\begin{aligned}
&\mathbb{P}(\tilde{Y}_{i2} \le y_2 | Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{P}(\tilde{Y}_{i2} \le y_2 | Y_{i0}, Y_{i1}, \gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i) | Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{P}(Y_{i2} \le y_2 | Y_{i0}, Y_{i1}, \gamma_i, \beta_i) | Y_{i0}, Y_{i1}) \\
&= \mathbb{P}(Y_{i2} \le y_2 | Y_{i0}, Y_{i1}),
\end{aligned}$$

which shows that (27) holds for the new process. Next, to show that the new data generating process also satisfies (28), note that in conditional expectations, conditioning on $(\gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1})$ is equivalent to conditioning on $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$ because $(\tilde{\gamma}_i, \tilde{\beta}_i)$ is a deterministic function of $(\gamma_i, \beta_i, Y_{i0}, Y_{i1})$. Then, by the law of iterated expectations, the following shows that the first line of (28) holds for the new data generating process:

$$\begin{aligned}
&\mathbb{E}(Y_{i1} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i0} | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i1} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i0} | \gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i1} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i0} | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i1} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i0} | \gamma_i, \beta_i, Y_{i0}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i1} - \gamma_i - \beta_i Y_{i0} + (Y_{i1} - Y_{i0}) k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \gamma_i, \beta_i, Y_{i0}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) \\
&= \mathbb{E}(0 + 0 | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}) = 0,
\end{aligned}$$

where the last line is obtained by the first line of (28) and by (29). Similarly, the following derivation shows that the second line of (28) holds for the new data generating process:

$$\begin{aligned}
&\mathbb{E}(\tilde{Y}_{i2} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i1} | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{E}(\tilde{Y}_{i2} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i1} | \gamma_i, \beta_i, \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{E}(\tilde{Y}_{i2} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i1} | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i2} - \tilde{\gamma}_i - \tilde{\beta}_i Y_{i1} | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i2} - \gamma_i - \beta_i Y_{i1} \\
&\qquad + Y_{i1} k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) - Y_{i1} k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) \\
&= \mathbb{E}(\mathbb{E}(Y_{i2} - \gamma_i - \beta_i Y_{i1} | \gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) = \mathbb{E}(0 | \tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}) = 0,
\end{aligned}$$

where the second-last equality follows from the second line of (28).

In summary, I have shown that, given the function k, $(\tilde{\gamma}_i, \tilde{\beta}_i, Y_{i0}, Y_{i1}, \tilde{Y}_{i2})$ is a new data generating process that satisfies (27) and (28). It now remains to specify the expression for k. Among many possible choices for k that satisfy (29), I consider the following expression:

$$k(r, b, y_0, y_1) = 1 - \frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i = r, \beta_i = b, Y_{i0} = y_0)}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i = r, \beta_i = b, Y_{i0} = y_0)} (y_1 - y_0).$$

The following calculation shows that this choice of k satisfies (29):

$$\begin{aligned}
&\mathbb{E}((Y_{i1} - Y_{i0}) k(\gamma_i, \beta_i, Y_{i0}, Y_{i1}) | \gamma_i, \beta_i, Y_{i0}) \\
&= \mathbb{E}\left((Y_{i1} - Y_{i0}) - \frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0})} (Y_{i1} - Y_{i0})^2 \middle| \gamma_i, \beta_i, Y_{i0} \right) \\
&= \mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0}) - \mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0}) = 0.
\end{aligned}$$

Then, under this choice of k, the expectation of $\tilde{\beta}_i$ is calculated as

$$\begin{aligned}
\mathbb{E}(\tilde{\beta}_i) &= \mathbb{E}(\beta_i + k(\gamma_i, \beta_i, Y_{i0}, Y_{i1})) \\
&= \mathbb{E}(\beta_i) + 1 - \mathbb{E}\left(\frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0})} (Y_{i1} - Y_{i0}) \right) \\
&= \mathbb{E}(\beta_i) + 1 - \mathbb{E}\left(\mathbb{E}\left(\frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0})} (Y_{i1} - Y_{i0}) \middle| \gamma_i, \beta_i, Y_{i0} \right) \right) \\
&= \mathbb{E}(\beta_i) + 1 - \mathbb{E}\left(\frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})^2}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0})} \right).
\end{aligned}$$

Note that, in the last term, the denominator is larger than the numerator because

$$\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0}) - \mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})^2 = \mathrm{Var}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0}) \geq 0,$$

where equality holds if $Y_{i1} - Y_{i0}$ has zero variance conditional on $(\gamma_i, \beta_i, Y_{i0})$, which is ruled out by the assumptions. Then, it follows that

$$\mathbb{E}(\tilde{\beta}_i) = \mathbb{E}(\beta_i) + 1 - \mathbb{E}\left(\frac{\mathbb{E}(Y_{i1} - Y_{i0} | \gamma_i, \beta_i, Y_{i0})^2}{\mathbb{E}((Y_{i1} - Y_{i0})^2 | \gamma_i, \beta_i, Y_{i0})} \right) > \mathbb{E}(\beta_i) + 1 - 1 = \mathbb{E}(\beta_i).$$

This implies that $\mathbb{E}(\tilde{\beta}_i)$ is strictly larger than $\mathbb{E}(\beta_i)$, which completes the proof. □

Proof of Lemma 1. Lemma 3 in Online Appendix B.3 implies that, if $\mathbb{E}(\beta_i)$ is point-identified, then

$$f^*(Y_{i0}, Y_{i1}, Y_{i2}) + g_1^*(\gamma_i, \beta_i, Y_{i0})\varepsilon_{i1} + g_2^*(\gamma_i, \beta_i, Y_{i0}, Y_{i1})\varepsilon_{i2} = \beta_i \tag{32}$$

almost surely for $(\gamma_i, \beta_i, Y_{i0}, Y_{i1}, Y_{i2}) \in \mathcal{C}$, where f^*, g_1^* and g_2^* are linear functionals on the spaces of finite and countably additive signed Borel measures that are absolutely continuous with respect to the Lebesgue measure. Then (32) implies that $S^*(Y_{i0}, Y_{i1}, Y_{i2}) = f^*(Y_{i0}, Y_{i1}, Y_{i2})$ since

$$\mathbb{E}(f^*(Y_{i0}, Y_{i1}, Y_{i2})|\beta_i) = \mathbb{E}\left(\beta_i - g_1^*(\gamma_i, \beta_i, Y_{i0})\varepsilon_{i1} - g_2^*(\gamma_i, \beta_i, Y_{i0}, Y_{i1})\varepsilon_{i2}|\beta_i\right) = \beta_i.$$

Conversely, if there exists $S^*(Y_{i0}, Y_{i1}, Y_{i2})$ satisfying $\mathbb{E}(S^*(Y_{i0}, Y_{i1}, Y_{i2})|\beta_i) = \beta_i$, then $\mathbb{E}(S^*(Y_{i0}, Y_{i1}, Y_{i2})) = \mathbb{E}(\mathbb{E}(S^*(Y_{i0}, Y_{i1}, Y_{i2})|\beta_i)) = \mathbb{E}(\beta_i)$, which completes the proof. □

Proof of Theorem 1. In essence, the proof follows the logic used to show the general result of Theorem 2. However, because Theorem 2 has not yet been introduced and Theorem 1 does not require the regularity assumption of Theorem 2, I present here a standalone proof of Theorem 1 under Assumptions 1 and 2.

Let $\lambda \in \mathbb{R}$ and let μ be a real vector that has the same dimension as R_{it}. Then, define the function

$$\mathcal{L}(\lambda, \mu, W_i, B_i) \equiv e'B_i + \mu' \sum_{t=1}^{T} R_{it}(Y_{it} - R_{it}'B_i) + \lambda \sum_{t=1}^{T} (R_{it}'B_i)(Y_{it} - R_{it}'B_i).$$

Note that $\mathbb{E}(\mathcal{L}) = \mathbb{E}(e'B_i)$ since

$$\mathbb{E}(\mathcal{L}) = \mathbb{E}(e'B_i) + \mu' \sum_{t=1}^{T} \mathbb{E}(R_{it}(Y_{it} - R_{it}'B_i)) + \lambda \sum_{t=1}^{T} \mathbb{E}((R_{it}'B_i)(Y_{it} - R_{it}'B_i))$$

and that the moment condition in (2) implies that

$$\mathbb{E}(R_{it}(Y_{it} - R_{it}'B_i)) = 0 \quad \text{and} \quad \mathbb{E}((R_{it}'B_i)(Y_{it} - R_{it}'B_i)) = 0.$$

Now, note that

$$\mathbb{E}(e'B_i) = \mathbb{E}(\mathcal{L}(\lambda, \mu, W_i, B_i)) \leq \mathbb{E}(\max_b \mathcal{L}(\lambda, \mu, W_i, b)),$$

since $\mathcal{L}(\lambda, \mu, W_i, B_i) \leq \max_b \mathcal{L}(\lambda, \mu, W_i, b)$ almost surely. Then, since this inequality holds for all (λ, μ), it follows that

$$\mathbb{E}(e'B_i) \leq \min_{\lambda,\mu} \mathbb{E}(\max_b \mathcal{L}(\lambda, \mu, W_i, b)),$$

which implies that the right-hand side is an upper bound of $\mathbb{E}(e'B_i)$. Similarly, it follows that

$$\mathbb{E}(e'B_i) = \mathbb{E}(\mathcal{L}(\lambda, \mu, W_i, B_i)) \geq \mathbb{E}(\min_b \mathcal{L}(\lambda, \mu, W_i, b)),$$

which then implies that

$$\mathbb{E}(e'B_i) \geq \max_{\lambda,\mu} \mathbb{E}(\min_b \mathcal{L}(\lambda, \mu, W_i, b)),$$

which implies that the right-hand side is a lower bound of $\mathbb{E}(e'B_i)$. Therefore, in summary, I obtain $L^* \leq \mathbb{E}(e'B_i) \leq U^*$ where

$$L^* \equiv \max_{\lambda,\ \mu} \mathbb{E} \left(\min_b \left[e'b + \mu' \sum_{t=1}^{T} R_{it}(Y_{it} - R_{it}'b) + \lambda \sum_{t=1}^{T} (R_{it}'b)(Y_{it} - R_{it}'b) \right] \right),$$

and

$$U^* \equiv \min_{\lambda,\ \mu} \mathbb{E} \left(\max_b \left[e'b + \mu' \sum_{t=1}^{T} R_{it}(Y_{it} - R_{it}'b) + \lambda \sum_{t=1}^{T} (R_{it}'b)(Y_{it} - R_{it}'b) \right] \right).$$

In what follows, I show that the closed-form expression for L^* and U^* coincide with those in Theorem 1. In the remainder of the proof, I focus on showing that the closed-form expression for U^* coincides with that in Theorem 1. The expression for L^* can be verified by a similar argument.

Now I derive the closed-form expression for U^*. Using the matrix notations R_i and Y_i, I can write U^* concisely as

$$U^* = \min_{\mu,\lambda} \mathbb{E} \left(\max_b \left[e'b + \mu' R_i' Y_i - \mu' R_i' R_i b + \lambda Y_i' R_i b - b'(\lambda R_i' R_i) b \right] \right).$$

Here, the inner maximization problem optimizes a quadratic polynomial in b, where $-\lambda R_i' R_i$ is the leading coefficient matrix. Note that if $\lambda > 0$, then by Assumption 2, the matrix $-\lambda R_i' R_i$ is negative definite, in which case this quadratic polynomial attains a finite closed-form maximum at the solution to the first-order condition with respect to b. On the other hand, if $\lambda \leq 0$, the polynomial's maximum diverges to $+\infty$ unless $R_i' R_i$ has a zero eigenvalue, which is ruled out by Assumption 2. Consequently, a finite upper bound is obtained only for $\lambda > 0$, and I can write U^* as:

$$U^* = \min_{\mu,\lambda>0} \mathbb{E}\left(\max_b \left[e'b + \mu' R_i' Y_i - \mu' R_i' R_i b + \lambda Y_i' R_i b - b'(\lambda R_i' R_i) b \right] \right).$$

Now, for $\lambda > 0$, I can solve for the closed-form maximum of the quadratic polynomial in b. The first-order condition with respect to b yields

$$e - R_i' R_i \mu + \lambda R_i' Y_i - 2(\lambda R_i' R_i) b = 0 \quad \Rightarrow \quad b^* = \frac{1}{2}(\lambda R_i' R_i)^{-1}(e + \lambda R_i' Y_i - R_i' R_i \mu).$$

Plugging this solution back into the expression for U^* yields:

$$\min_{\lambda>0,\ \mu} \mathbb{E}\left(\mu' R_i' Y_i + \frac{1}{4\lambda} \left[e + \lambda R_i' Y_i - R_i' R_i \mu \right]' (R_i' R_i)^{-1} \left[e + \lambda R_i' Y_i - R_i' R_i \mu \right] \right). \tag{33}$$

I solve this problem with respect to μ for a fixed λ. The first-order condition with respect to μ given λ is:

$$\mathbb{E}(R_i' Y_i) + \frac{1}{2\lambda}\mathbb{E}(R_i' R_i)\mu - \frac{1}{2\lambda} e - \frac{1}{2}\mathbb{E}(R_i' Y_i) = 0.$$

The optimal μ that solves this first-order condition is $\mu^* = \mathbb{E}(R_i' R_i)^{-1}[e - \lambda \mathbb{E}(R_i' Y_i)]$. I substitute this into (33) and then solve the resulting expression with respect to λ. The first-order condition with respect to λ is:

$$\frac{1}{\lambda^2}\left[e'\mathbb{E}((R_i' R_i)^{-1}) e - e'\mathbb{E}(R_i' R_i)^{-1} e \right] = \mathbb{E}(R_i' Y_i' (R_i' R_i)^{-1} R_i' Y_i) - \mathbb{E}(R_i' Y_i)' \mathbb{E}(R_i' R_i)^{-1} \mathbb{E}(R_i' Y_i).$$

Since $\lambda > 0$, the optimal λ that satisfies this first-order condition is:

$$\lambda^* = \sqrt{\frac{e'\mathbb{E}((R_i' R_i)^{-1}) e - e'\mathbb{E}(R_i' R_i)^{-1} e}{\mathbb{E}(R_i' Y_i' (R_i' R_i)^{-1} R_i' Y_i) - \mathbb{E}(R_i' Y_i)' \mathbb{E}(R_i' R_i)^{-1} \mathbb{E}(R_i' Y_i)}}. \tag{34}$$

Substituting (34) back into (33) yields the expression for $\tilde{U}$ in Theorem 1.

The numerator and denominator in (34) are both weakly positive, and each is equal to zero if and only if $(R_i' R_i)^{-1} e$ and $(R_i' R_i)^{-1} R_i' Y_i$ are degenerate across individuals, respectively. To show this, one can apply the following proposition to the functions $E(R_i' R_i) = e'(R_i' R_i)^{-1} e$ and $D(R_i' Y_i, R_i' R_i) = (R_i' Y_i)' (R_i' R_i)^{-1} R_i' Y_i$. □

Proposition 8 (Kiefer, 1959, Lemma 3.2). *For an integer $l > 0$, let $A_1, \ldots, A_l$ be $n \times m$ matrices and $B_1, \ldots, B_l$ be nonsingular positive definite and symmetric $n \times n$ matrices. Let $a_1, \ldots, a_l$*

be positive real numbers such that $\sum_k a_k = 1$. Then

$$\sum_{k=1}^{l} a_k A_k' B_k^{-1} A_k - \left[\sum_{k=1}^{l} a_k A_k\right]' \left[\sum_{k=1}^{l} a_k B_k\right]^{-1} \left[\sum_{k=1}^{l} a_k A_k\right] \geq 0$$

where '$\geq$' is the partial ordering defined in terms of positive semidefinite matrices. In addition, the equality holds if and only if $B_1^{-1}A_1 = \ldots = B_l^{-1}A_l$.

Proof of Proposition 2. I first show that $\mathcal{V}_S$ is invertible, for which I show that $\mathcal{V}_S$ is positive definite, i.e.,

$$x' S_i R_i (R_i' R_i)^{-1} R_i' S_i' x > 0$$

with positive probability for every nonzero x.

Note first that Assumption 3 implies $S_i' x \neq 0$ with positive probability for every nonzero x. Note also that Assumption 2 implies that R_i has full column rank, meaning that $y' R_i (R_i' R_i)^{-1} R_i' y > 0$ for any nonzero vector y. Now, for every nonzero x, define $y = S_i' x$. Then $y \neq 0$ with positive probability, and it follows that

$$x' S_i R_i (R_i' R_i)^{-1} R_i' S_i' x = y' R_i (R_i' R_i)^{-1} R_i' y > 0$$

with positive probability. This proves that $\mathcal{V}_S$ is positive definite, and thus invertible.

Now I prove the remainder of Proposition 2. Following the proof of Theorem 1, let $\lambda \in \mathbb{R}$ and let μ_t be a real vector that has the same dimension as S_{it}, and define the function

$$\begin{aligned}\mathcal{L} &\equiv e' B_i + \sum_{t=1}^{T} \mu_t' S_{it} (Y_{it} - R_{it}' B_i) + \lambda \sum_{t=1}^{T} (R_{it}' B_i)(Y_{it} - R_{it}' B_i) \\ &= e' B_i + \mu' S_i (Y_i - R_i B_i) + \lambda B_i' R_i' (Y_i - R_i B_i),\end{aligned}$$

where $\mu = (\mu_1', \mu_2', \ldots, \mu_T')$ is a real vector that attaches all μ_t vectors into a single vector. Then, following the proof of Theorem 1, $L_S \leq \mathbb{E}(e' B_i) \leq U_S$ where

$$L_S \equiv \max_{\mu,\lambda} \mathbb{E}\left(\min_b \left[e' b + \mu' S_i (Y_i - R_i b) + \lambda b' R_i' (Y_i - R_i b)\right]\right),$$

and

$$U_S \equiv \min_{\mu,\lambda} \mathbb{E}\left(\max_b \left[e' b + \mu' S_i (Y_i - R_i b) + \lambda b' R_i' (Y_i - R_i b)\right]\right).$$

In what follows, I show that the closed-form expression for L_S and U_S coincide with those in Proposition 2. In the remainder of the proof, I focus on showing that the closed-form expression for U_S coincides with that in Proposition 2. The expression for L_S can be verified by a similar argument.

Now I derive the closed-form expression for U_S. I first rewrite U_S as

$$U_S = \min_{\mu,\lambda} \mathbb{E}\left(\max_b \left[\mu' S_i Y_i + (e + \lambda R_i' Y_i - R_i' S_i' \mu)' b - b' \left(\lambda R_i' R_i\right) b\right] \right).$$

The inner maximization problem of U_S optimizes a quadratic polynomial in b, where $-\lambda R_i' R_i$ is the leading coefficient matrix. Note that if $\lambda > 0$, then by Assumption 2, this matrix is negative definite, in which case the quadratic polynomial attains a finite closed-form maximum. On the other hand, if $\lambda \leq 0$, the polynomial's maximum diverges to $+\infty$ unless $R_i' R_i$ has a zero eigenvalue, which is ruled out by Assumption 2. Consequently, a finite upper bound is obtained only for $\lambda > 0$, and I can write U_S as:

$$U_S = \min_{\mu,\lambda>0} \mathbb{E}\left(\max_b \left[\mu' S_i Y_i + (e + \lambda R_i' Y_i - R_i' S_i' \mu)' b - b' \left(\lambda R_i' R_i\right) b\right] \right).$$

Now, for $\lambda > 0$, I can solve for the closed-form maximum of the quadratic polynomial in b. The first order condition with respect to b yields

$$e + \lambda R_i' Y_i - R_i' S_i' \mu - 2(\lambda R_i' R_i) b = 0 \quad \Rightarrow \quad b^* = \frac{1}{2}(\lambda R_i' R_i)^{-1}\left(e + \lambda R_i' Y_i - R_i' S_i' \mu\right).$$

Plugging this solution back into the expression for U_S yields:

$$\min_{\lambda>0,\ \mu} \mathbb{E}\left(\mu' S_i Y_i + \frac{1}{4\lambda}\left[e + \lambda R_i' Y_i - R_i' S_i' \mu\right]' (R_i' R_i)^{-1} \left[e + \lambda R_i' Y_i - R_i' S_i' \mu\right] \right). \tag{35}$$

I solve this problem with respect to μ for a fixed λ. The first-order condition with respect to μ given λ is:

$$\mathbb{E}(S_i Y_i) - \frac{1}{2\lambda}\mathbb{E}\left(S_i R_i (R_i' R_i)^{-1} \left[e + \lambda R_i' Y_i - R_i' S_i' \mu\right]\right) = 0.$$

The optimal μ that solves this first-order condition is

$$\mu^* = \mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' S_i')^{-1} \left[\mathbb{E}(S_i R_i (R_i' R_i)^{-1}) e + \lambda \left(\mathbb{E}(S_i R_i (R_i' R_i)^{-1} R_i' Y_i) - 2\mathbb{E}(S_i Y_i)\right)\right]$$

which I can write concisely as, using the notation from the main text:

$$\mu^* = \mathcal{V}_S^{-1}\left[\mathcal{P}_S e + \lambda\left(\mathcal{Y}_S - 2Y_S\right)\right].$$

I now substitute this into (35). First, expand (35) and obtain

$$\min_{\lambda>0,\ \mu} \Big\{ \mu' Y_S + \frac{1}{4\lambda} \Big[e'\mathbb{E}((R_i'R_i)^{-1})e + \lambda^2 m_0 + \mu' \mathcal{V}_S \mu \\ + 2\lambda e'\mathbb{E}(\widehat{B}_i) - 2e'\mathcal{P}_S' \mu - 2\lambda \mathcal{Y}_S' \mu \Big] \Big\},$$

where I used the notation from the main text to write it concisely. I then substitute the expression for μ^* into the above, obtaining:

$$\min_{\lambda>0} \Big\{ Y_S' \mathcal{V}_S^{-1} \left[\mathcal{P}_S e + \lambda \left(\mathcal{Y}_S - 2Y_S \right) \right] \\ + \frac{1}{4\lambda} \Big[e'\mathbb{E}((R_i'R_i)^{-1})e + \lambda^2 m_0 + \left[\mathcal{P}_S e + \lambda \left(\mathcal{Y}_S - 2Y_S \right) \right]' \mathcal{V}_S^{-1} \left[\mathcal{P}_S e + \lambda \left(\mathcal{Y}_S - 2Y_S \right) \right] \\ + 2\lambda e'\mathbb{E}(\widehat{B}_i) - 2e'\mathcal{P}_S' \mathcal{V}_S^{-1} \left[\mathcal{P}_S e + \lambda \left(\mathcal{Y}_S - 2Y_S \right) \right] - 2\lambda \mathcal{Y}_S' \mathcal{V}_S^{-1} \left[\mathcal{P}_S e + \lambda \left(\mathcal{Y}_S - 2Y_S \right) \right] \Big] \Big\}.$$

Expanding this expression and collecting terms with respect to λ yields the expression:

$$\min_{\lambda>0} \Big\{ \frac{1}{4\lambda} [e'\mathbb{E}((R_i'R_i)^{-1})e - e'\mathcal{P}_S' \mathcal{V}_S^{-1} \mathcal{P}_S e] + \frac{\lambda}{4} [m_0 - (\mathcal{Y}_S - 2Y_S)' \mathcal{V}_S^{-1} (\mathcal{Y}_S - 2Y_S)] \\ + \frac{1}{2} [2e'\mathcal{P}_S' \mathcal{V}_S^{-1} Y_S + e'\mathbb{E}(\widehat{B}_i) - e'\mathcal{P}_S' \mathcal{V}_S^{-1} \mathcal{Y}_S] \Big\}.$$

The first order condition with respect to λ then yields

$$\lambda^* = \sqrt{\frac{e'\mathbb{E}((R_i'R_i)^{-1})e - e'\mathcal{P}_S' \mathcal{V}_S^{-1} \mathcal{P}_S e}{m_0 - (\mathcal{Y}_S - 2Y_S)' \mathcal{V}_S^{-1} (\mathcal{Y}_S - 2Y_S)}}.$$

Substituting this expression yields the expression for U_S in Proposition 2. □

Proof of Proposition 3. I first show that $\mathcal{V}_M - M_0$ is invertible. Note that

$$\mathcal{V}_M - M_0 = \mathbb{E}(M_i' R_i (R_i'R_i)^{-1} R_i' M_i) - \mathbb{E}(M_i' M_i) = \mathbb{E}(M_i' (R_i (R_i'R_i)^{-1} R_i' - I) M_i).$$

Since $R_i(R_i'R_i)^{-1}R_i'$ is a projection matrix, $R_i(R_i'R_i)^{-1}R_i' - I$ is negative semidefinite. In addition, Assumption 5 implies that $R_i'R_i$ is positive definite and that M_i is not in the column space of R_i with positive probability. These imply that $M_i'(R_i(R_i'R_i)^{-1}R_i' - I)M_i$ is negative semidefinite and that it is negative definite with positive probability. Therefore, its expectation $\mathcal{V}_M - M_0$ is negative definite and thus invertible.

Next, I show that $\mathcal{V}$ is invertible. First, as shown in the proof of Proposition 2, $\mathcal{V}_S$ is positive definite under Assumptions 5 and 6. I have also shown that $\mathcal{V}_M - M_0$ is negative definite, which implies that $-(\mathcal{V}_M - M_0)$ is positive definite. Therefore, the quantity $\mathcal{V}_S - (C - \mathcal{C})(\mathcal{V}_M - M_0)(C - \mathcal{C})'$ is positive definite and thus invertible.

Now I prove the remainder of Proposition 3. Following the proof of Proposition 2, let $\lambda \in \mathbb{R}$ and let μ_t be a real vector that has the same dimension as S_{it}. For a fixed value of δ, define the function

$$\begin{aligned}\mathcal{L}(\delta) &\equiv e'B_i + \sum_{t=1}^{T} \mu_t' S_{it}(Y_{it} - R_{it}'B_i - M_{it}'\delta) + \lambda \sum_{t=1}^{T} (R_{it}'B_i + M_{it}'\delta)(Y_{it} - R_{it}'B_i - M_{it}'\delta) \\ &= e'B_i + \mu' S_i(Y_i - R_iB_i - M_i\delta) + \lambda(B_i'R_i' + \delta'M_i')(Y_i - R_iB_i - M_i\delta),\end{aligned}$$

where $\mu = (\mu_1', \mu_2', \ldots, \mu_T')$ is a real vector that attaches all μ_t vectors into a single vector. Then, following the proof of Theorem 1, $L_M(\delta) \le \mathbb{E}(e'B_i) \le U_M(\delta)$ where

$$L_M(\delta) \equiv \max_{\mu,\lambda} \mathbb{E}\left(\min_b \left[e'b + \mu' S_i(Y_i - R_ib - M_i\delta) + \lambda(b'R_i' + \delta'M_i')(Y_i - R_ib - M_i\delta) \right] \right),$$

and

$$U_M(\delta) \equiv \min_{\mu,\lambda} \mathbb{E}\left(\max_b \left[e'b + \mu' S_i(Y_i - R_ib - M_i\delta) + \lambda(b'R_i' + \delta'M_i')(Y_i - R_ib - M_i\delta) \right] \right).$$

Then, an outer bound of $\mathbb{E}(e'B_i)$ is obtained by taking the union of $[L_M(\delta), U_M(\delta)]$ over all $\delta \in \mathbb{R}^{q_m+p_m}$ (rather than over all admissible δ only). In other words, it follows that $L_M \le \mathbb{E}(e'B_i) \le U_M$ where

$$L_M = \min_{\delta \in \mathbb{R}^{q_m+p_m}} L_M(\delta), \quad \text{and} \quad U_M = \max_{\delta \in \mathbb{R}^{q_m+p_m}} U_M(\delta).$$

In what follows, I show that the closed-form expression for L_M and U_M coincide with those in Proposition 3. In the remainder of the proof, I focus on showing that the closed-form expression for U_M coincides with that in Proposition 3. The expression for L_S can be verified by a similar argument.

Now I derive the closed-form expression for U_M, for which I first derive the closed-form expression for $U_M(\delta)$. For a fixed δ, the inner maximization problem of $U_M(\delta)$ optimizes a quadratic polynomial in b, where $-\lambda R_i'R_i$ is the leading coefficient matrix. Note that if $\lambda > 0$, then by Assumption 5, this matrix is negative definite, in which case the quadratic polynomial attains a finite closed-form maximum. On the other hand, if $\lambda \le 0$, the polynomial's maximum diverges to $+\infty$ unless $R_i'R_i$ has a zero eigenvalue, which is ruled out by Assumption 5. Consequently, a finite upper bound is obtained only for $\lambda > 0$. Then, for $\lambda > 0$, I solve for the first order condition with respect to b and substitute the solution back into $U_M(\delta)$, which yields

$$\begin{aligned} U_M(\delta) = \min_{\lambda>0,\ \mu} \mathbb{E}\Big(& \mu' S_i Y_i - \mu' S_i M_i \delta + \lambda\delta' M_i' Y_i - \lambda\delta' M_i' M_i \delta \\ & + \frac{1}{4\lambda}\left[e - R_i' S_i' \mu + \lambda R_i' Y_i - 2\lambda R_i' M_i \delta\right]' (R_i' R_i)^{-1} \left[e - R_i' S_i' \mu + \lambda R_i' Y_i - 2\lambda R_i' M_i \delta\right] \Big). \end{aligned} \tag{36}$$

I then expand the terms in (36), which yields the expression

$$\begin{aligned} U_M(\delta) = \min_{\lambda>0,\ \mu} \Big\{ & \mu' Y_S - \mu' C\delta + \lambda\delta' Y_M - \lambda\delta' M_0 \delta \\ & + \frac{1}{4\lambda} e' \mathbb{E}((R_i' R_i)^{-1}) e + \frac{\lambda}{4} m_0 + \frac{1}{4\lambda} \mu' \mathcal{V}_S \mu + \lambda\delta' \mathcal{V}_M \delta \\ & + \frac{1}{2} e' \mathbb{E}(\widehat{B}_i) - \frac{1}{2\lambda} e' \mathcal{P}_S' \mu - \frac{1}{2} \mathcal{Y}_S' \mu - \delta' \mathcal{P}_M e - \lambda\delta' \mathcal{Y}_M + \mu' \mathcal{C}\delta \Big\}. \end{aligned}$$

Let $L(\lambda, \mu, \delta)$ be the objective function of the above, so that $U_M(\delta) = \min_{\lambda>0,\ \mu} L(\lambda, \mu, \delta)$. Then I can write

$$U_M = \max_{\delta} U_M(\delta) = \max_{\delta} \min_{\lambda>0,\ \mu} L(\lambda, \mu, \delta).$$

I note two properties about $L(\lambda, \mu, \delta)$. First, it can be shown that $L(\lambda, \mu, \delta)$ is convex in (λ, μ), which I state later as Proposition 4. Second, $L(\lambda, \mu, \delta)$ is a quadratic polynomial in δ, where the leading coefficient matrix is $\lambda(\mathcal{V}_M - M_0)$. Since $\lambda > 0$ and $\mathcal{V}_M - M_0$ is negative definite, this implies that $L(\lambda, \mu, \delta)$ is concave in δ. These two properties then imply:

$$U_M = \max_{\delta} \min_{\lambda>0,\ \mu} L(\lambda, \mu, \delta) = \min_{\lambda>0,\ \mu} \max_{\delta} L(\lambda, \mu, \delta).$$

Now I solve for the latter expression. I first solve $L(\lambda, \mu, \delta)$ with respect to δ for a fixed (λ, μ). The optimal δ that solves the first order condition is

$$\delta^* = -\frac{1}{2\lambda}(\mathcal{V}_M - M_0)^{-1}\left(\lambda(Y_M - \mathcal{Y}_M) - C'\mu - \mathcal{P}_M e + \mathcal{C}'\mu\right).$$

I then substitute this into $L(\lambda, \mu, \delta)$, which yields the following expression for U_M:

$$\begin{aligned} \min_{\lambda>0,\ \mu} \Big\{ & \mu' Y_S + \frac{1}{4\lambda} e' \mathbb{E}((R_i' R_i)^{-1}) e + \frac{\lambda}{4} m_0 + \frac{1}{4\lambda} \mu' \mathcal{V}_S \mu + \frac{1}{2} e' \mathbb{E}(\widehat{B}_i) - \frac{1}{2\lambda} e' \mathcal{P}_S' \mu - \frac{1}{2} \mathcal{Y}_S' \mu \\ & \frac{1}{4\lambda}\left(\lambda(Y_M - \mathcal{Y}_M) - \mathcal{C}'\mu - \mathcal{P}_M e + \mathcal{C}'\mu\right)' (\mathcal{V}_M - M_0)^{-1} \left(\lambda(Y_M - \mathcal{Y}_M) - \mathcal{C}'\mu - \mathcal{P}_M e + \mathcal{C}'\mu\right) \Big\}. \end{aligned}$$

Next, I solve this expression with respect to μ for a fixed $\lambda > 0$. The optimal μ that solves the first order condition is

$$\mu^* = \mathcal{V}^{-1}\left[\mathcal{P}_S e + \lambda \mathcal{Y}_S - 2\lambda Y_S + (C - \mathcal{C})(\mathcal{V}_M - M_0)^{-1}(\mathcal{P}_M e + \lambda \mathcal{Y}_M - \lambda Y_M)\right].$$

I substitute this expression back into the above expression for U_M, obtaining:

$$U_M = \min_{\lambda > 0}\left(\frac{\lambda}{4}\mathcal{D}_M + \mathcal{B}_M + \frac{1}{4\lambda}\mathcal{E}_M\right),$$

where $\mathcal{D}_M$, $\mathcal{B}_M$, and $\mathcal{E}_M$ are as defined in Proposition 3. The first order condition with respect to λ then yields

$$\lambda^* = \sqrt{\frac{\mathcal{E}_M}{\mathcal{D}_M}}.$$

Substituting this expression yields the expression for U_M in Proposition 3. □

Proof of Theorem 2. In what follows, I show that (14) is the dual representation of (13). The proof is a direct application of the duality theorem for linear programming over topological vector spaces (Anderson, 1983). The same argument applies to (15).

To apply the theorem, I first rewrite (13) into a standard form of linear programming, for which I introduce additional notation. Recall that $\mathcal{M}_{W \times B}$ is a linear space of finite and countably additive signed Borel measures on $\mathcal{W} \times \mathcal{B}$. Let $\overline{\mathcal{F}}_{W \times B}$ be the dual space of $\mathcal{M}_{W \times B}$, and let $\mathcal{F}_{W \times B}$ be the space of all bounded Borel measurable functions on $\mathcal{W} \times \mathcal{B}$. Note that $\mathcal{F}_{W \times B}$ is a linear subspace of $\overline{\mathcal{F}}_{W \times B}$.

For $P \in \mathcal{M}_{W \times B}$ and $f \in \overline{\mathcal{F}}_{W \times B}$, define the *dual pairing*

$$\langle P, f \rangle = \int f dP.$$

Similarly, let $\mathcal{M}_W$ be the linear space of finite and countably additive signed Borel measures on $\mathcal{W}$. Let $\overline{\mathcal{F}}_W$ be the dual space of $\mathcal{M}_W$, and let $\mathcal{F}_W$ be the space of all bounded Borel measurable functions on $\mathcal{W}$. Note that $\mathcal{F}_W$ is a linear subspace of $\overline{\mathcal{F}}_W$. In addition, define $\mathcal{G} = \mathbb{R}^K \times \mathcal{M}_W$ and $\mathcal{H} = \mathbb{R}^K \times \overline{\mathcal{F}}_W$, and let $g = (g_1, \ldots, g_K, P_g)$ and $h = (\lambda_1, \ldots, \lambda_K, f_h)$ be their generic elements. Note that $\mathcal{H}$ is the dual space of $\mathcal{G}$. Define the dual pairing

$$\langle g, h \rangle = \sum_{k=1}^{K} \lambda_k g_k + \int f_h dP_g.$$

Next, define a linear map $A : \mathcal{M}_{W\times B} \mapsto \mathcal{G}$ by

$$A(P) = \left(\int \phi_1 dP, \ldots, \int \phi_K dP, P(\cdot, \mathcal{B}) \right).$$

Since each ϕ_k is bounded, A is a bounded (thus continuous) linear operator. Moreover, note that

$$\langle A(P), h \rangle = \sum_{k=1}^{K} \lambda_k \int \phi_k dP + \int_{\mathcal{W}} f_h(w) P(dw, \mathcal{B}).$$

It is straightforward to show that

$$\int_{\mathcal{W}} f_h(w) P(dw, \mathcal{B}) = \int_{\mathcal{W}\times\mathcal{B}} f_h(w) dP(w, v).$$

Then I obtain

$$\langle A(P), h \rangle = \sum_{k=1}^{K} \lambda_k \int \phi_k dP + \int f_h dP = \int \left[\sum_{k=1}^{K} \lambda_k \phi_k + f_h \right] dP \equiv \langle P, A^*(h) \rangle, \tag{37}$$

where $A^*(h) : \mathcal{H} \mapsto \overline{\mathcal{F}}_{W\times B}$ is defined as

$$A^*(h) = \sum_{k=1}^{K} \lambda_k \phi_k + f_h.$$

Equation (37) shows that A^* is the adjoint of A. With these notations, I rewrite (13) as a standard form of linear programming:

$$\min_{P \in \mathcal{M}_{W\times V}} \langle P, m \rangle \qquad \text{subject to} \qquad A(P) = c, \qquad P \geq 0, \tag{38}$$

where $c = (0, \ldots, 0, P_W)$. I now apply the duality theorem to (38), where I verify the conditions for the duality theorem later in the proof. The dual problem of (38) is given by:

$$\max_{h \in \mathcal{H}} \langle c, h \rangle \qquad \text{subject to} \qquad m - A^*(h) \geq 0,$$

which I can write more concretely as:

$$\max_{\lambda_1, \ldots, \lambda_K \in \mathbb{R},\ f_h \in \overline{\mathcal{F}}_W} \int f_h dP_W \qquad \text{subject to} \qquad \sum_{k=1}^{K} \lambda_k \phi_k + f_h \leq m. \tag{39}$$

I now simplify (39) further. I rearrange the constraint of (39):

$$f_h(w) \leq m(w, b) - \sum_{k=1}^{K} \lambda_k \phi_k(w, b).$$

The left-hand side does not involve v. Therefore:

$$f_h(w) \leq \min_{b \in \mathcal{B}} \left[m(w, b) - \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right] \quad \text{for all} \quad w \in \mathcal{W}.$$

Since (39) maximizes the expectation of f_h, the optimal solution f_h^* for a fixed $(\lambda_1, \ldots, \lambda_K)$ is given by:

$$f_h^*(w) = \min_{b \in \mathcal{B}} \left[m(w, b) - \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right] \tag{40}$$

almost surely in P_W. If not, i.e., if $f_h^*(w)$ is strictly less than the right-hand side of (40) with positive probability in P_W, then one can increase the value of the objective by increasing f_h^* on a set of positive probability. I then substitute (40) into (39), which yields:

$$\max_{\lambda_1, \ldots, \lambda_K \in \mathbb{R}} \int \min_{b \in \mathcal{B}} \left[m(w, b) - \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right] dP_W(w).$$

The above display remains equivalent even if the signs of $(\lambda_1, \ldots, \lambda_K)$ are switched, because they are choice variables supported on $\mathbb{R}^K$. Switching the signs of $(\lambda_1, \ldots, \lambda_K)$ then gives:

$$\max_{\lambda_1, \ldots, \lambda_K \in \mathbb{R}} \int \min_{b \in \mathcal{B}} \left[m(w, b) + \sum_{k=1}^{K} \lambda_k \phi_k(w, b) \right] dP_W(w) \tag{41}$$

which is exactly the expression in (14).

It remains to verify that the conditions for the duality theorem holds, so that the optimal value of the primal problem in (38) equals to that of the dual problem in (41). Note that, by assumption, the dual problem in (41) attains a finite maximum value, which I denote by L. Then, since the set $\{P \in \mathcal{M}_{W \times V} \mid P \geq 0\}$ is closed in $\mathcal{M}_{W \times V}$, the sufficient conditions for Theorem 4 in Anderson (1983) are satisfied. This theorem then implies that the primal problem in (38) attains its minimum value at L, and thus strong duality holds. Note that Theorem 4 in Anderson (1983) requires that the dual problem has a sequence of choice variables $(\lambda_1, \ldots, \lambda_K)$ whose objective function values converge to the optimal value L, which trivially holds if the optimal value L is attained. □

Proof of Proposition 4. This proof shows that G_L is concave in λ. The proof for the convexity of G_U is similar. Let $\lambda_1 = (\lambda_{11}, \ldots, \lambda_{1K})$ and $\lambda_2 = (\lambda_{21}, \ldots, \lambda_{2K})$ be two arbitrary points in $\mathbb{R}^K$. Then, for any $t \in [0, 1]$ and $w \in \mathcal{W}$:

$$\begin{aligned}
&G_L(t\lambda_1 + (1-t)\lambda_2, w) \\
&= \min_{b \in \mathcal{B}} \left\{ t \left[m(w,b) + \sum_{k=1}^{K} \lambda_{1k}\phi_k(w,b) \right] + (1-t) \left[m(w,b) + \sum_{k=1}^{K} \lambda_{2k}\phi_k(w,b) \right] \right\} \\
&\geq t \min_{b \in \mathcal{B}} \left\{ m(w,b) + \sum_{k=1}^{K} \lambda_{1k}\phi_k(w,b) \right\} + (1-t) \min_{b \in \mathcal{B}} \left\{ m(w,b) + \sum_{k=1}^{K} \lambda_{2k}\phi_k(w,b) \right\} \\
&= tG_L(\lambda_1, w) + (1-t)G_L(\lambda_2, w),
\end{aligned}$$

which is the definition of concavity. □

Proof of Lemma 2 As stated in (39) in the proof of Theorem 2, the sharp lower bound of θ is given by

$$\max_{\lambda_1, \ldots, \lambda_K \in \mathbb{R},\, f_h \in \overline{\mathcal{F}}_W} \int f_h dP_W \qquad \text{subject to} \qquad \sum_{k=1}^{K} \lambda_k \phi_k + f_h \leq m \tag{42}$$

where all notations follow the proof of Theorem 2. Similarly, the sharp upper bound of θ is given by

$$\min_{\lambda_1, \ldots, \lambda_K \in \mathbb{R},\, f_h \in \overline{\mathcal{F}}_W} \int f_h dP_W \qquad \text{subject to} \qquad \sum_{k=1}^{K} \lambda_k \phi_k + f_h \geq m. \tag{43}$$

Suppose, by the way of contradiction, that θ is point-identified but there do not exist $S^* \in \overline{\mathcal{F}}_W$ and $\lambda_1^*, \ldots, \lambda_K^* \in \mathbb{R}$ such that, almost surely:

$$\sum_{k=1}^{K} \lambda_k^* \phi_k + S^* = m.$$

Then the optimal solution $(\lambda_1^l, \ldots, \lambda_K^l, S^l)$ to (42) must satisfy its constraint $\sum_{k=1}^{K} \lambda_k^l \phi_k + S^l \leq m$ with strict inequality on a set of positive Lebesgue measure on $\mathcal{W} \times \mathcal{B}$. Similarly, the optimal solution $(\lambda_1^u, \ldots, \lambda_K^u, S^u)$ to (43) must satisfy its constraint $\sum_{k=1}^{K} \lambda_k^u \phi_k + S^u \geq m$ with strict inequality on a set of positive Lebesgue measure on $\mathcal{W} \times \mathcal{B}$. Then it follows that:

$$\mathbb{E}(S^l) = \mathbb{E}\left(\sum_{k=1}^{K} \lambda_k^l \phi_k + S^l \right) < \mathbb{E}(m) < \mathbb{E}\left(\sum_{k=1}^{K} \lambda_k^u \phi_k + S^u \right) = \mathbb{E}(S^u)$$

where strict inequalities follow because the density of (W_i, B_i) is strictly positive. This implies that the sharp lower bound $\mathbb{E}(S^l)$ is strictly less than the sharp upper bound $\mathbb{E}(S^u)$, which is a contradiction since θ is assumed to be point-identified.

Conversely, suppose there exists $(S^*, \lambda_1^*, \ldots, \lambda_K^*)$ such that $\sum_{k=1}^K \lambda_k^* \phi_k + S^* = m$. Then:

$$\mathbb{E}(S^*) = \mathbb{E}\left(\sum_{k=1}^K \lambda_k \phi_k + S^*\right) = \mathbb{E}(m) = \theta,$$

which shows that θ is point-identified by $\mathbb{E}(S^*)$. $\square$

B Online Appendix: Extensions and Discussions

B.1 Extension to multivariate random coefficient models

Results from this paper extend to a multivariate version of (1), a system of random coefficient models:

$$\mathbf{Y}_{it} = \mathbf{Z}_{it}'\gamma_i + \mathbf{X}_{it}'\beta_i + \mathbf{e}_{it},$$

where $\mathbf{Y}_{it}$ is a $D \times 1$ vector of dependent variables, $\mathbf{Z}_{it}$ is a $D \times q$ matrix of strictly exogenous regressors, $\mathbf{X}_{it}$ is a $D \times p$ matrix of sequentially exogenous regressors, and $\mathbf{e}_{it}$ is a $D \times 1$ vector of idiosyncratic error terms. Assume that

$$\mathbb{E}(\mathbf{e}_{it} | \gamma_i, \beta_i, \mathbf{Z}_i, \mathbf{X}_i^t) = 0,$$

which is a multivariate extension of (2). The following is an example of a multivariate random coefficient model.

Example 4 (Joint model of household earnings and consumption behavior)**.** One can combine (3) and (4) to construct a joint lifecycle model of earnings and consumption behavior. In particular, when the time t consumption equation is combined with the time $t+1$ earnings equation, a system of random coefficient models is obtained:

$$\begin{aligned} C_{it} &= \gamma_{i1} + \gamma_{i2} Y_{it} + \beta_{i1} A_{it} + \nu_{it}, \\ Y_{i,t+1} &= \gamma_{i3} + \beta_{i2} Y_{it} + \varepsilon_{it}, \end{aligned}$$

which can be written in the following matrix form:

$$\begin{pmatrix} C_{it} \\ Y_{i,t+1} \end{pmatrix} = \begin{pmatrix} 1 & Y_{it} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_{i1} \\ \gamma_{i2} \\ \gamma_{i3} \end{pmatrix} + \begin{pmatrix} A_{it} & 0 \\ 0 & Y_{it} \end{pmatrix} \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \end{pmatrix} + \begin{pmatrix} \nu_{it} \\ \varepsilon_{it} \end{pmatrix}.$$

In this model, the γ_is and β_is can freely correlate among themselves and with (Y_{i0}, A_{i1}), allowing for correlation between earnings and consumption processes.

B.2 Alternative proof of Proposition 1

This proof is an application of the general result in Online Appendix B.3. Suppose that the regularity conditions stated as Assumption 13 in Online Appendix B.3 hold. Also, for notational simplicity, suppose that $\mathcal{C} = \mathcal{C}_0^5$, where $\mathcal{C}_0$ is a compact subset of $\mathbb{R}$. The proof can be easily modified for a generic compact set $\mathcal{C}$.

Assume that $\mathbb{E}(\beta_i)$ is point-identified, from which I derive a contradiction. Lemma 3 in Online Appendix B.3 implies that, if $\mathbb{E}(\beta_i)$ is point-identified, then:

$$f^*(Y_{i0}, Y_{i1}, Y_{i2}) + g_1^*(\gamma_i, \beta_i, Y_{i0})\varepsilon_{i1} + g_2^*(\gamma_i, \beta_i, Y_{i0}, Y_{i1})\varepsilon_{i2} = \beta_i \tag{44}$$

almost surely in $(\gamma_i, \beta_i, Y_{i0}, Y_{i1}, Y_{i2})$, where $f^* : \mathcal{C}_0^3 \mapsto \mathbb{R}$, $g_1^* : \mathcal{C}_0^3 \mapsto \mathbb{R}$ and $g_2^* : \mathcal{C}_0^4 \mapsto \mathbb{R}$ are linear functionals on the spaces of finite and countably additive signed Borel measures that are absolutely continuous with respect to the Lebesgue measure. Substituting $\varepsilon_{it} = Y_{it} - \gamma_i - \beta_i Y_{i,t-1}$ in (44) yields, almost surely in $(\gamma_i, \beta_i, Y_{i0}, Y_{i1}, Y_{i2})$,

$$f^*(Y_{i0}, Y_{i1}, Y_{i2}) + g_1^*(\gamma_i, \beta_i, Y_{i0})(Y_{i1} - \gamma_i - \beta_i Y_{i0}) + g_2^*(\gamma_i, \beta_i, Y_{i0}, Y_{i1})(Y_{i2} - \gamma_i - \beta_i Y_{i1}) = \beta_i. \tag{45}$$

Now, consider any $\gamma, \tilde{\gamma}, \beta, y_0, y_1, y_2 \in \mathcal{C}_0$ such that $\gamma \neq \tilde{\gamma}$. I then evaluate (45) at $(\gamma_i, \beta_i, Y_{i0}, Y_{i1}, Y_{i2}) = (\gamma, \beta, y_0, y_1, y_2)$ and at $(\tilde{\gamma}, \beta, y_0, y_1, y_2)$ and take the difference between the two. This yields:

$$\begin{aligned} &(y_1 - \tilde{\gamma} - \beta y_0)\triangle_{\tilde{\gamma},\gamma} g_1^* - (\tilde{\gamma} - \gamma) g_1^*(\gamma, \beta, y_0) \\ &+ (y_2 - \tilde{\gamma} - \beta y_1)\triangle_{\tilde{\gamma},\gamma} g_2^* - (\tilde{\gamma} - \gamma) g_2^*(\gamma, \beta, y_0, y_1) = 0 \end{aligned} \tag{46}$$

where $\triangle_{\tilde{\gamma},\gamma} g_1^* \equiv g_1^*(\tilde{\gamma}, \beta, y_0) - g_1^*(\gamma, \beta, y_0)$ and $\triangle_{\tilde{\gamma},\gamma} g_2^* \equiv g_2^*(\tilde{\gamma}, \beta, y_0, y_1) - g_2^*(\gamma, \beta, y_0, y_1)$.

In (46), note that the variable y_2 appears only in the third term. In addition, (46) must hold almost surely for all $\gamma, \tilde{\gamma}, \beta, y_0, y_1, y_2 \in \mathcal{C}_0$ such that $\gamma \neq \tilde{\gamma}$, and in particular for any choice of $y_2 \in \mathcal{C}_0$. This implies that the third term must not depend on y_2, which implies that, almost surely:

$$\triangle_{\tilde{\gamma},\gamma} g_2^* = 0,$$

which means that g_2^* is almost surely a constant function on γ:

$$g_2^*(\gamma, \beta, y_0, y_1) = g_2^*(\beta, y_0, y_1). \tag{47}$$

If not, i.e., if $\triangle_{\tilde{\gamma},\gamma} g_2^* \neq 0$ on a subset of $\mathcal{C}_0^6$ with positive Lebesgue measure, one can change the value of y_2 without altering $(\gamma, \tilde{\gamma}, \beta, y_0, y_1)$ within that subset to violate (46) on a set of a positive measure.

Next, consider any $\gamma, \beta, \tilde{\beta}, y_0, y_1, y_2 \in \mathcal{C}_0$ such that $\beta \neq \tilde{\beta}$. I then evaluate (45) at $(\gamma, \beta, y_0, y_1, y_2)$ and $(\gamma, \tilde{\beta}, y_0, y_1, y_2)$ and take the difference between the two. This yields:

$$\begin{aligned}&(y_1 - \gamma - \tilde{\beta} y_0) \triangle_{\tilde{\beta},\beta} g_1^* - (\tilde{\beta} - \beta) y_0 g_1^*(\gamma, \beta, y_0) \\ &+ (y_2 - \gamma - \tilde{\beta} y_1) \triangle_{\tilde{\beta},\beta} g_2^* - (\tilde{\beta} - \beta) y_1 g_2^*(\beta, y_0, y_1) = \tilde{\beta} - \beta\end{aligned} \tag{48}$$

where $\triangle_{\tilde{\beta},\beta} g_1^* \equiv g_1^*(\gamma, \tilde{\beta}, y_0) - g_1^*(\gamma, \beta, y_0)$ and $\triangle_{\tilde{\beta},\beta} g_2^* = g_2^*(\tilde{\beta}, y_0, y_1) - g_2^*(\beta, y_0, y_1)$. In (48), note that y_2 appears only in the third term. This implies $g_2^*(\beta, y_0, y_1) = g_2^*(y_0, y_1)$ almost surely, similarly to the argument leading to (47). Then (46) simplifies to:

$$(y_1 - \tilde{\gamma} - \beta y_0) \triangle_{\tilde{\gamma},\gamma} g_1^* - (\tilde{\gamma} - \gamma) g_1^*(\gamma, \beta, y_0) - (\tilde{\gamma} - \gamma) g_2^*(y_0, y_1) = 0. \tag{49}$$

Let $\gamma, \tilde{\gamma}, \hat{\gamma} \in \mathcal{C}_0$ be such that $\hat{\gamma} - \tilde{\gamma} = \tilde{\gamma} - \gamma$. I then evaluate (49) at $(\gamma, \tilde{\gamma}, \beta, y_0, y_1)$ and $(\tilde{\gamma}, \hat{\gamma}, \beta, y_0, y_1)$ and take the difference between the two. This yields:

$$(y_1 - \hat{\gamma} - \beta y_0) \left(\triangle_{\hat{\gamma},\tilde{\gamma}} g_1^* - \triangle_{\tilde{\gamma},\gamma} g_1^* \right) - (\hat{\gamma} - \tilde{\gamma}) \triangle_{\tilde{\gamma},\gamma} g_1^* - (\tilde{\gamma} - \gamma) \triangle_{\tilde{\gamma},\gamma} g_1^* = 0. \tag{50}$$

In (50), note that y_1 appears only in the first term, which implies $\triangle_{\hat{\gamma},\tilde{\gamma}} g_1^* - \triangle_{\tilde{\gamma},\gamma} g_1^* = 0$ almost surely, similarly to the argument leading to (47). Then (50) simplifies to:

$$(\hat{\gamma} - \tilde{\gamma}) \triangle_{\tilde{\gamma},\gamma} g_1^* + (\tilde{\gamma} - \gamma) \triangle_{\tilde{\gamma},\gamma} g_1^* = 0, \tag{51}$$

which implies $\triangle_{\tilde{\gamma},\gamma} g_1^* = 0$ since $\hat{\gamma} - \tilde{\gamma} = \tilde{\gamma} - \gamma \neq 0$. This implies that g_1^* is almost surely a constant function over γ, i.e., $g_1^*(\gamma, \beta, y_0) = g_1^*(\beta, y_0)$. Then (48) simplifies to:

$$\begin{aligned}&(y_1 - \gamma - \tilde{\beta} y_0) \triangle_{\tilde{\beta},\beta} g_1^* - (\tilde{\beta} - \beta) y_0 g_1^*(\beta, y_0) \\ &+ (y_2 - \gamma - \tilde{\beta} y_1) \triangle_{\tilde{\beta},\beta} g_2^* - (\tilde{\beta} - \beta) y_1 g_2^*(y_0, y_1) = \tilde{\beta} - \beta.\end{aligned} \tag{52}$$

Let $\beta, \tilde{\beta}, \hat{\beta} \in \mathcal{C}_0$ be such that $\hat{\beta} - \tilde{\beta} = \tilde{\beta} - \beta$. Evaluating (52) at $(\gamma, \hat{\beta}, \tilde{\beta}, y_0, y_1, y_2)$ and at $(\gamma, \tilde{\beta}, \beta, y_0, y_1, y_2)$ and taking the difference yields $g_1^*(\beta, y_0) = g_1^*(y_0)$, similarly to the argument leading to $g_1^*(\gamma, \beta, y_0) = g_1^*(\beta, y_0)$ from (50). Then (45) simplifies to:

$$f^*(y_0, y_1, y_2) + g_1^*(y_0)(y_1 - \gamma - \beta y_0) + g_2^*(y_0, y_1)(y_2 - \gamma - \beta y_1) = \beta$$

almost surely for all $(\gamma, \beta, y_0, y_1, y_2)$. This is a linear identity in (γ, β), so the coefficients of γ and β must match on both sides of the equation. In particular, equating the coefficients on γ implies that $-g_1^* - g_2^* = 0$, and equating the coefficients on β implies $-y_0 g_1^* - y_1 g_2^* = 1$. Solving these two equations for (g_1^*, g_2^*) yields, almost surely:

$$g_1^* = \frac{1}{y_1 - y_0}, \quad g_2^* = \frac{-1}{y_1 - y_0}.$$

However, g_1^* cannot be a function of y_1, which is a contradiction. □

B.3 Identification under conditional moment restrictions

This subsection studies moment equality models that incorporate both conditional and unconditional moment restrictions. Consider the following extension of Assumption 7.

Assumption 12. The random vectors (W_i, B_i) satisfy:

$$\begin{aligned} \mathbb{E}(\phi_k(W_i, B_i)) &= 0, \quad k = 1, \ldots, K_U, \\ \mathbb{E}(\psi_k(W_i, B_i) | A_{ik}) &= 0, \quad k = 1, \ldots, K_C, \end{aligned}$$

where $\phi_k, \psi_k : \mathcal{W} \times \mathcal{B} \mapsto \mathbb{R}$ are moment functions, $A_{i1}, \ldots, A_{iK_C}$ are subvectors of (W_i, B_i), and $K_U, K_C \in \mathbb{N}$ are the number of unconditional and conditional moment restrictions, respectively.

Under Assumption 12, I characterize the identified set of

$$\theta = \mathbb{E}(m(W_i, B_i))$$

for some known function $m : \mathcal{W} \times \mathcal{B} \mapsto \mathbb{R}$. For brevity of notation, let A'_{ik} be the subvector of (W_i, B_i) collecting the variables not included in A_{ik} so that (A_{ik}, A'_{ik}) is a rearrangement of (W_i, B_i). Accordingly, any function $f(w, b)$ on $\mathcal{W} \times \mathcal{B}$ can be equivalently written as $f(a_k, a'_k)$ on $\mathcal{A}_k \times \mathcal{A}'_k$, where $\mathcal{A}_k \times \mathcal{A}'_k$ is the rearrangement of $\mathcal{W} \times \mathcal{B}$ according to (A_{ik}, A'_{ik}). I assume the following regularity conditions.

Assumption 13. The following conditions hold.

(i) $\mathcal{W} \times \mathcal{B}$ is a compact set in a Euclidean space.

(ii) $(m, \phi_1, \ldots, \phi_{K_U}, \psi_1, \ldots, \psi_{K_C})$ are L^∞ with respect to the Lebesgue measure.

(iii) The distribution of (W_i, B_i) is absolutely continuous with respect to the Lebesgue measure, and its density p is L^∞ with respect to the Lebesgue measure.

(iv) There exists a joint density p_0 of (W_i, B_i) such that it satisfies Assumption 12, its marginal density with respect to W_i equals to the observed density of W_i, and it is strictly positive on $\mathcal{W} \times \mathcal{B}$.

Assumption 13 (i) and (ii) are similar to Assumption 8. Assumption 13 (iii) and (iv) are restrictive, but they are useful enough for Lemma 1 and an alternative proof of Proposition 1 in Online Appendix B.2.

Under these assumptions, I obtain the following theorem and lemma, which are counterparts of Theorem 2 and Lemma 2, respectively, by characterizing the identified set I of θ and providing a necessary and sufficient condition for point identification of θ.

Theorem 3. *Suppose that either of these two conditions hold:*

(A) Assumptions 12 and 13 (i)-(iii) hold, and the optimization problems

$$L = \max_{\{\lambda_k\}_{k=1}^{K_U}, \{\mu_k\}_{k=1}^{K_C}} \mathbb{E}\left[\min_{b \in \mathcal{B}} \left\{ m(W_i, b) + \sum_{k=1}^{K_U} \lambda_k \phi_k(W_i, b) + \sum_{k=1}^{K_C} \mu_k(A_k(W_i, b))\psi_k(W_i, b) \right\}\right] \tag{53}$$

and

$$U = \min_{\{\lambda_k\}_{k=1}^{K_U}, \{\mu_k\}_{k=1}^{K_C}} \mathbb{E}\left[\max_{b \in \mathcal{B}} \left\{ m(W_i, b) + \sum_{k=1}^{K_U} \lambda_k \phi_k(W_i, b) + \sum_{k=1}^{K_C} \mu_k(A_k(W_i, b))\psi_k(W_i, b) \right\}\right] \tag{54}$$

possess finite solutions, where $\lambda_k \in \mathbb{R}$ for $k = 1, \ldots, K_U$ and $\mu_k : \mathcal{A}_k \mapsto \mathbb{R}$ for $k = 1, \ldots, K_C$, and $A_k(w, b)$ denotes the value of A_{ik} given $W_i = w$ and $B_i = b$.

(B) Assumptions 12 and 13 (i)-(iv) hold.

If either (A) or (B) holds, then $I = [L, U]$.

Proof. The proof focuses on showing (53). The same argument applies to (54).

Let $\mathcal{M}_{W \times B}$ be the space of finite and countably additive signed Borel measures that are absolutely continuous with respect to the Lebesgue measure. Using absolute continuity, I identify each element of $\mathcal{M}_{W\times B}$ by its density $p : \mathcal{W} \times \mathcal{B} \mapsto \mathbb{R}$. Let p_W be the density of the observed data distribution P_W. Then, the identified set I is defined by

$$
\begin{aligned}
I \equiv \Bigg\{ \int m(w,b)p(w,b)\,d(w,b) \,\Bigg|\, & p \in \mathcal{M}_{W\times B}, \quad p \geq 0, \\
& \int \phi_k(w,b)p(w,b)d(w,b) = 0, \quad k = 1,\dots,K_U, \\
& \int \psi_k(a_k,a_k')p(a_k,a_k')da_k' = 0 \text{ for all } a_k \in \mathcal{A}_k, \quad k = 1,\dots,K_C, \\
& \int p(w,b)db = p_W(w) \ \text{ for all } w \in \mathcal{W} \Bigg\},
\end{aligned}
$$

where a_k is an element of $\mathcal{A}_k$ and a_k' is an element of $\mathcal{A}_k'$. The second line above represents the unconditional moment restrictions, while the third line represents the conditional moment restrictions.

The lower bound of I is then given by the infinite-dimensional linear program

$$
\begin{aligned}
\min_{p\in\mathcal{M}_{W\times B},\ p\geq 0} \int m(w,b)p(w,b)d(w,b) \qquad & \text{subject to} \\
\int \phi_k(w,b)p(w,b)d(w,b) = 0, \quad & k = 1,\dots,K_U, \\
\int \psi_k(a_k,a_k')p(a_k,a_k')da_k' = 0, \text{ for all } a_k \in \mathcal{A}_k, \quad & k = 1,\dots,K_C, \\
\int p(w,b)db = p_W(w) \ \text{ for all } w \in \mathcal{W}. &
\end{aligned} \tag{55}
$$

Now I show that (53) is the dual representation of (55), by directly applying the duality theorem of linear programming for topological vector spaces (Anderson, 1983), for which I introduce additional notation. Let $L^2(\mathcal{W}\times\mathcal{B})$ be the space of all L^2 functions on $\mathcal{W}\times\mathcal{B}$, and let $L^2(\mathcal{W})$ be the space of all L^2 functions on $\mathcal{W}$. I also let $L^2(\mathcal{A}_k)$ be the space of all L^2 functions on $\mathcal{A}_k$.

Define $\mathcal{G}$ and $\mathcal{H}$ as $\mathcal{G} = \mathcal{H} = \mathbb{R}^K \times L^2(\mathcal{A}_1) \times \ldots \times L^2(\mathcal{A}_{K_C}) \times L^2(\mathcal{W})$. I denote their generic elements as $g = (g_1,\dots,g_{K_U},\overline{g}_1,\dots,\overline{g}_{K_C},f_g)$ and $h = (\lambda_1,\dots,\lambda_{K_U},\mu_1,\dots,\mu_{K_C},f_h)$, respectively. Note that $\mathcal{H}$ is a dual space of $\mathcal{G}$.

Define a linear map $A : \mathcal{M}_{W\times B} \mapsto \mathcal{G}$ by

$$
A(p) = \left(\int \phi_1 p \ d(w,b), \dots, \int \phi_K p \ d(w,b), \int \psi_k p \ da_1', \dots, \int \psi_k p \ da_{K_C}', \int p \ db \right).
$$

The map A is a bounded (thus continuous) linear operator because the functions ϕ_k and

ψ_k are assumed to be bounded. Next, define the dual pairing as

$$\langle A(P), h\rangle = \sum_{k=1}^{K_U} \lambda_k \int \phi_k p \, d(w, vb) + \sum_{k=1}^{K_C} \iint \psi_k p \, da'_k \, \mu_k da_k + \int f_h \int pdbdw.$$

It is straightforward to show that

$$\iint \psi_k p \, da'_k \, \mu_k da_k = \int \psi_k \mu_k p \, d(w, b)$$

and

$$\int f_h \int pdbdw = \int f_h p \, d(w, b).$$

Then, I obtain

$$\langle A(P), h\rangle = \int \left[\sum_{k=1}^{K_U} \lambda_k \phi_k + \sum_{k=1}^{K_C} \mu_k \psi_k + f_h \right] p(w, b) d(w, b) \equiv \langle p, A^*(h)\rangle, \tag{56}$$

where $A^*(h) : \mathcal{H} \mapsto L^2(\mathcal{W} \times \mathcal{B})$ is defined as

$$A^*(h) = \sum_{k=1}^{K_U} \lambda_k \phi_k + \sum_{k=1}^{K_C} \mu_k \psi_k + f_h.$$

Equation (56) shows that A^* is the adjoint of A.

Then, similar to the proof of Theorem 2, I apply the strong duality theorem of linear programming for topological vector spaces to (55). Under the conditions of (A), I apply Theorem 4 of Anderson (1983), similarly to the proof of Theorem 2. Under the conditions of (B), I use Theorem 9 of Anderson (1983). The sufficient conditions for this Theorem 9 are satisfied as follows. First, Assumption 13 (iv) verifies the condition that "there is x_0 in the interior of P with $Ax_0 = b$" in Theorem 9 of Anderson (1983). Second, Assumption 13 (i)-(iii) ensures that the primal problem in (55) possesses a finite solution, which verifies the condition that "EP has finite value" in Theorem 9 of Anderson (1983). In addition, its implicit condition that A is continuous is also satisfied.

Consequently, under either conditions of (A) or (B), the strong duality holds. Therefore, the optimal solution to (55) is equal to the solution to

$$\max_{\lambda_1,\ldots,\lambda_{K_U},\mu_1,\ldots,\mu_{K_C},f_h} \int f_h(w) p_w(w) dw \qquad \text{subject to} \qquad \sum_{k=1}^{K_U} \lambda_k \phi_k + \sum_{k=1}^{K_C} \mu_k \psi_k + f_h \leq m. \tag{57}$$

Then, simplifying (57) as in the proof of Theorem 2 yields the expression in (53).

□

Lemma 3. *Suppose that the conditions (B) of Theorem 3 hold. Then θ is point-identified if and only if there exists a function S^* which is a linear functional on $\mathcal{M}_W$ (which is the projection of $\mathcal{M}_{W\times B}$ onto $\mathcal{W}$), real numbers $\lambda_1^*, \dots, \lambda_K^* \in \mathbb{R}$, and functions $\mu_1^*, \dots, \mu_K^*$ which are $L^2(\mathcal{A}_1), \dots, L^2(\mathcal{A}_{K_C})$ functions, respectively, such that:*

$$m(W_i, b) + \sum_{k=1}^{K_U} \lambda_k \phi_k(W_i, b) + \sum_{k=1}^{K_C} \mu_k(A_k(W_i, b))\psi_k(W_i, b) = S^*(W_i)$$

almost surely on $\mathcal{W} \times \mathcal{B}$. When such S^* exists, θ is identified by $\theta = \mathbb{E}(S^*(W_i))$.

Proof. As in (57) in the proof of Theorem 3, the sharp lower bound of θ is given by

$$\max_{\lambda_1,\dots,\lambda_{K_U},\mu_1,\dots,\mu_{K_C},f_h} \int f_h(w)p_w(w)dw \qquad \text{subject to} \qquad \sum_{k=1}^{K_U} \lambda_k\phi_k + \sum_{k=1}^{K_C} \mu_k\psi_k + f_h \le m,$$

where all notation follows the proof of Theorem 3. Similarly, the sharp upper bound of θ is given by

$$\min_{\lambda_1,\dots,\lambda_{K_U},\mu_1,\dots,\mu_{K_C},f_h} \int f_h(w)p_w(w)dw \qquad \text{subject to} \qquad \sum_{k=1}^{K_U} \lambda_k\phi_k + \sum_{k=1}^{K_C} \mu_k\psi_k + f_h \ge m.$$

Lemma 3 then follows by replicating the argument in the proof of Lemma 2.

□

B.4 Identified set for a general variance parameter

In this subsection, I consider identification of a general variance parameter. Recall the dynamic random coefficient model defined in (1) and (2):

$$Y_{it} = R_{it}' B_i + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it}|B_i, Z_i, X_i^t) = 0, \qquad t = 1, \dots, T,$$

where $R_{it} = (Z_{it}', X_{it}')'$. I consider the second moments of the random coefficients:

$$V_e = \mathbb{E}(e_1' B_i B_i' e_2)$$

where e_1 and e_2 are real-valued vectors that the researcher chooses. For example, if $e_1 = e_2 = (1, 0, \dots, 0)'$, then V_e is the second moment of the first entry of B_i.

The key idea for identifying the mean in Section 3 is to consider a moment condition that is quadratic in B_i so that it can "dominate" the linear term $e'B_i$. By the same idea, to "dominate" the term $e_1'B_iB_i'e_2$, I consider a moment condition that is fourth order in B_i. Specifically, I consider the moment restrictions

$$\mathbb{E}\left(\sum_{t=1}^{T}(R_{it}'B_i)^3\varepsilon_{it}\right) = 0, \quad \text{and} \quad \mathbb{E}\left(S_{it}\varepsilon_{it}\right) = 0 \quad \text{for} \quad t = 1, \dots, T, \tag{58}$$

where I replaced the first moment restriction in (11) with a fourth order restriction in B_i. A direct application of Theorem 2 then yields the following bounds for V_e, which I state without proof.

Proposition 9. *Suppose that Assumptions 1 to 3 and 8 hold, and that Equation (58) holds. Then* $L_V \leq V_e \leq U_V$ *where*

$$L_V = \max_{\lambda>0,\ \mu\in\mathbb{R}^L} \mathbb{E}\left[\min_{b\in\mathcal{B}}\left\{e_1'bb'e_2 + \lambda\sum_{t=1}^{T}(R_{it}'b)^3(Y_{it} - R_{it}'b) + \mu'S_i(Y_i - R_ib)\right\}\right],$$

and

$$U_V = \min_{\lambda<0,\ \mu\in\mathbb{R}^L} \mathbb{E}\left[\max_{b\in\mathcal{B}}\left\{e_1'bb'e_2 + \lambda\sum_{t=1}^{T}(R_{it}'b)^3(Y_{it} - R_{it}'b) + \mu'S_i(Y_i - R_ib)\right\}\right].$$

In Proposition 9, the optimization over b requires global optimization of a fourth-order polynomial, for which no closed-form solution exists. Instead, this problem can be solved numerically using the semidefinite relaxation approach (Lasserre, 2010, 2015), which transforms the polynomial optimization problem into a convex optimization problem over a variable that is a semidefinite matrix.

B.5 Overidentification in the general inference procedure

In this subsection, I discuss a simple heuristic modification of the general inference procedure in Section 5.2 that practically deals with the overidentification, assuming that the model is correctly specified. Note that this heuristic method may yield spuriously narrow confidence intervals for θ even if the model is partially identified. This occurs because the procedure does not formally account for overidentification or misspecification as in Stoye (2020) and Andrews and Kwon (2024). Nonetheless, I check the performance of the heuristic procedure via simulation at the end of this subsection, and I find that it achieves reasonable coverage rates.

I first note that the inference procedure of Andrews and Shi (2017) is still implementable under overidentification, since $T_{AS}(\theta)$ and $c^{(b)}_{AS,GMS}(\theta)$ can still be computed for each fixed θ. The practical difficulty, however, is that the finite-sample optimizers $\hat{\lambda}^L_N$ and $\hat{\lambda}^U_N$ diverge, making it hard to search for the supremum in the neighborhoods of $\hat{\lambda}^L_N$ and $\hat{\lambda}^U_N$. To resolve this issue, I propose to compute the finite-sample optimizers with L^1 penalties:

$$\begin{aligned}\tilde{\lambda}^L_N(\zeta) &= \max_{\lambda\in\mathbb{R}^K}\left[\frac{1}{N}\sum_{i=1}^N G_L(\lambda, W_i) - \zeta\sum_{k\in K_0}|\lambda_k|\right],\\ \tilde{\lambda}^U_N(\zeta) &= \min_{\lambda\in\mathbb{R}^K}\left[\frac{1}{N}\sum_{i=1}^N G_U(\lambda, W_i) + \zeta\sum_{k\in K_0}|\lambda_k|\right],\end{aligned} \tag{59}$$

where $\zeta > 0$ is the penalty parameter and $K_0 \subseteq \{1,\ldots,K\}$ is the index set of the penalized moment restrictions. In practice, one may simply take $K_0 = \{1,\ldots,K\}$.

This modification has the following econometric interpretation. For $\zeta \geq 0$, consider the following relaxation of the moment conditions:

$$\mathbb{E}(\phi_k(W_i, B_i)) = 0 \text{ for all } k \notin K_0 \quad \text{and} \quad |\mathbb{E}(\phi_k(W_i, B_i))| \leq \zeta \text{ for all } k \in K_0, \tag{60}$$

which reduces to Assumption 7 if $\zeta = 0$. This is similar in spirit to the minimally relaxed identified set in Andrews and Kwon (2024). The following proposition then characterizes the smallest $\zeta \geq 0$ for which the finite sample satisfies the relaxed moment restrictions in (60), therefore resolving the overidentification issue.

Proposition 10. *Given the sample $(W_1,\ldots,W_N)$, consider the linear program*

$$\begin{aligned}\min_{P\in\mathcal{M}_{W\times B},\, P\geq 0,\, \zeta\geq 0}\zeta \qquad \textit{subject to} \qquad & \left|\int \phi_k(W_i, B_i)dP\right| \leq \zeta, \quad k\in K_0,\\ & \int \phi_k(W_i, B_i)dP = 0, \quad k\notin K_0,\\ & \int P(w, dB_i) = \hat{P}_W(w) \ \textit{for all } w\in\mathcal{W},\end{aligned} \tag{61}$$

where $\hat{P}_W$ is the empirical distribution of W_i constructed from $(W_1,\ldots,W_N)$. Then, under Assumption 8, its solution equals to the solution to:

$$\max_{\lambda\in\mathbb{R}^K}\frac{1}{N}\sum_{i=1}^N\min_{b\in\mathcal{B}}\left\{\sum_{k=1}^K\lambda_k\phi_k(W_i, b)\right\} \qquad \textit{subject to} \qquad \sum_{k\in K_0}|\lambda_k| \leq 1. \tag{62}$$

Proof. I can rewrite (61) as:

$$\min_{P\in\mathcal{M}_{W\times B},\ P\geq 0,\ \zeta\geq 0}\zeta \qquad \text{subject to} \qquad \begin{aligned} &\int \phi_k(W_i,B_i)dP \leq \zeta, \quad k\in K_0,\\ &\int \phi_k(W_i,B_i)dP \geq -\zeta, \quad k\in K_0,\\ &\int \phi_k(W_i,B_i)dP = 0, \quad k\notin K_0,\\ &\int P(w,dB_i) = \hat{P}_W(w) \ \text{ for all } w\in\mathcal{W}, \end{aligned}$$

Then, similarly to the proof of Theorem 2, I can invoke the dualty theorem with inequality constraints (Anderson, 1983) to obtain (62) as the simplified dual representation of (61). □

Let ζ^* be the solution to (62). Then, for any $\zeta \geq \zeta^*$, it follows that the estimated bounds for θ under the ζ-relaxed moment restrictions in (60) are given by the L^1-penalized optimizations. The following proposition shows this result for the lower bound. The result for the upper bound holds similarly.

Proposition 11. *Given the sample $(W_1,\ldots,W_N)$ and given $\zeta\in\mathbb{R}$, consider the linear program that finds the smallest value of $\theta=\mathbb{E}(m(W_i,B_i))$ that satisfies (60):*

$$\min_{P\in\mathcal{M}_{W\times B},\ P\geq 0}\int m(W_i,B_i)dP \qquad \textit{subject to} \qquad \begin{aligned} &\left|\int \phi_k(W_i,B_i)dP\right| \leq \zeta, \quad k\in K_0,\\ &\int \phi_k(W_i,B_i)dP = 0, \quad k\notin K_0,\\ &\int P(w,db) = \hat{P}_W(w) \ \textit{ for all } w\in\mathcal{W}, \end{aligned} \tag{63}$$

where $\hat{P}_W$ is the empirical distribution of W_i constructed from $(W_1,\ldots,W_N)$. Then, under Assumption 8, its solution equals to

$$\tilde{L} = \max_{\lambda\in\mathbb{R}^K}\left[\frac{1}{N}\sum_{i=1}^{N} G_L(\lambda,W_i) - \zeta\sum_{k\in K_0}|\lambda_k|\right]. \tag{64}$$

Proof. I can rewrite (63) as:

$$\min_{P \in \mathcal{M}_{W\times B},\ P\geq 0} \int m(W_i, B_i) dP \qquad \text{subject to} \qquad \begin{aligned} &\int \phi_k(W_i, B_i) dP \leq \zeta, \quad k \in K_0, \\ &\int \phi_k(W_i, B_i) dP \geq -\zeta, \quad k \in K_0, \\ &\int \phi_k(W_i, B_i) dP = 0, \quad k \notin K_0, \\ &\int P(w, db) = \hat{P}_W(w) \ \text{ for all } w \in \mathcal{W}. \end{aligned}$$

Then, similarly to the proof of Theorem 2, I can invoke the dualty theorem with inequality constraints (Anderson, 1983) to obtain (64) as the simplified dual representation of (63). □

Proposition 11 implies that the L^1-penalized finite sample optimizers $\tilde{\lambda}_N^L(\zeta)$ and $\tilde{\lambda}_N^U(\zeta)$ defined in (59) are precisely the maximizer in (64) for the lower bound and and the corresponding minimizer for the upper bound. I then implement the procedure of Andrews and Shi (2017) with a modification that I restrict the supremum to be calculated over the neighborhoods of $\tilde{\lambda}_N^L(\zeta)$ and $\tilde{\lambda}_N^U(\zeta)$ rather than the entire Q^K space of λ. Note that, if the supremum were taken over all of Q^K, both $T_{AS}(\theta)$ and $c_{AS,GMS}^{(b)}(\theta)$ would all diverge. As overidentification is resolved as $N \to \infty$, the supremum over Q^K becomes finite and the original procedure of Andrews and Shi (2017) can be applied.

In what follows, I check the performance of this heuristic method by simulation. Consider a data generating process

$$Y_{it} = \alpha_i + \beta_i Y_{i,t-1} + \varepsilon_{it}, \quad t = 1, \ldots, T,$$

with $T = 10$, where $\alpha_i \sim Unif[-3, 3]$, $\beta_i \sim Unif[0, 1]$, and $\varepsilon_{it} \sim N(0, 1)$ are independent random variables. I generate the initial value by $Y_{i0} = \beta_i + z_i$, where $z_i \sim N(0, 1)$ is another independent random variable. I construct the confidence interval for $\mathbb{E}(\beta_i)$ using the moment conditions in (11), where I set $S_{it} = \{1, Y_{i,\max\{0,t-5\}}, \ldots, Y_{i,t-1}\}$ and relax the first moment restriction $\mathbb{E}\left(\sum_{t=1}^T (R_{it}' B_i)\varepsilon_{it}\right) = 0$, which means that I take $K_0 = \{1\}$. To circumvent the computational difficulty in calcuating the population identified set of this model, I generate $100,000$ observations from this model and treat them as a finite population, for which the population identified set can be calculated using the bounds in Proposition 2.

Under this setup, I implement the heuristic modification of Andrews and Shi (2017) where I calculate the supremum over a finite grid of L points in the neighborhoods of

$\tilde{\lambda}_N^L(\zeta)$ and $\tilde{\lambda}_N^U(\zeta)$. I obtain this grid by adding Gaussian perturbations to $\tilde{\lambda}_N^L(\zeta)$ and $\tilde{\lambda}_N^U(\zeta)$, while including these optimizers themselves. I then obtain the critical values via 100 bootstrap replications, for a nominal coverage rate of 0.95. The simulated coverage rate is obtained by 1000 Monte Carlo replications.

Table 4 reports the simulated coverage rates for various sample sizes N and the grid sizes L. The results show that, as long as L is sufficiently large, the heuristic confidence interval achieves reasonable coverage rates.

	$L = 50$	$L = 100$	$L = 200$
$N = 500$	0.918	0.916	0.941
$N = 750$	0.952	0.947	0.959

Table 4: Simulated coverage rates for the heuristic inference procedure, where the supremum is evaluated with a total of L finite number of points in the neighborhoods of $\tilde{\lambda}_N^L(\zeta)$ and $\tilde{\lambda}_N^U(\zeta)$. The nominal coverage probability is 0.95.

B.6 Estimations with transitory income processes

In this subsection, I examine robustness of the results in Section 6 to existence of a transitory income process. In particular, I examine if the confidence intervals for $\mathbb{E}(\rho_i)$ change if there is a transitory income process added to (24).

To estimate the confidence intervals in the presence of a transitory income process, I consider the model $Y_{it} = \tilde{Y}_{it} + \varepsilon_{it}$, where Y_{it} is the raw log-earnings data, $\tilde{Y}_{it}$ is the log-earnings without the transitory income process, and ε_{it} is an i.i.d. Gaussian transitory income process. Note that the models in the main text do not involve transitory income processes, which corresponds to $\varepsilon_{it} = 0$.

I conduct estimation and inference for $\mathbb{E}(\rho_i)$ in the presence of ε_{it} in a two-step procedure. First, I use the approach Arellano and Bonhomme (2021) to numerically recover pseudo-observations of $\tilde{Y}_{it}$ when Y_{it} is observed and the distribution of ε_{it} is known. Second, I apply estimation procedures for the RIP-RC, HIP-RC, RIP-RC-J, and HIP-RC-J models to the numerically recovered pseudo-observations of $\tilde{Y}_{it}$. The estimation results from this procedure are presented in Table 5. The estimation results are qualitatively similar to those in Table 2 in the main text. In particular, the upper confidence limits of $\mathbb{E}(\rho_i)$ are significantly less than 1, and the confidence intervals for the RIP and the HIP processes show substantial overlap.

In what follows, I describe in detail the procedure that I used to numerically recover $\tilde{Y}_{it}$ from the model $Y_{it} = \tilde{Y}_{it} + \varepsilon_{it}$. I first describe the method proposed in Arellano and

Parameter	RIP-RC	HIP-RC	RIP-RC-J	HIP-RC-J
$\mathbb{E}(\rho_i)$	[0.475, 0.644]	[0.319, 0.626]	[0.549, 0.677]	[0.352, 0.631]

Table 5: Confidence intervals of $\mathbb{E}(\rho_i)$ for the RIP type and the HIP type processes with heterogeneous coefficients, after obtaining pseudo-observations of $\tilde{Y}_{it}$ without the transitory income process using the method of Arellano and Bonhomme (2021). The nominal coverage probability is 0.95. These confidence intervals are robust to overidentification and model misspecification.

Bonhomme (2021). They considered a model

$$Z = X + \varepsilon \tag{65}$$

where all variables are scalar[7] and X is independent of ε. In this model, Z is observed, but X and ε are not observed. Instead, the distribution of ε is known. The objective of Arellano and Bonhomme (2021) is to obtain pseudo-observations of X, given the observations of Z and the knowledge on the distribution of ε.

Let $\mathbb{P}_Z$, $\mathbb{P}_X$ and $\mathbb{P}_\varepsilon$ be the probability distributions of Z, X and ε, respectively. Let $\mathbb{P}_{X+\varepsilon}$ be the distribution of $X + \varepsilon$, which is equal to the convolution of $\mathbb{P}_X$ and $\mathbb{P}_\varepsilon$. The second-order Wasserstein distance between Z and $X + \varepsilon$, denoted by $W_2(\mathbb{P}_Z, \mathbb{P}_{X+\varepsilon})$, is defined by:

$$W_2(\mathbb{P}_Z, \mathbb{P}_{X+\varepsilon}) = \left(\min_{\pi \in \Pi(\mathbb{P}_Z, \mathbb{P}_{X+\varepsilon})} \int ||z - \hat{z}||^2 d\pi(z, \hat{z}) \right)^{1/2}, \tag{66}$$

where $\Pi(\mathbb{P}_Z, \mathbb{P}_{X+\varepsilon})$ is the set of couplings of $\mathbb{P}_Z$ and $\mathbb{P}_{X+\varepsilon}$, i.e., joint distributions of Z and $X + \varepsilon$ whose marginal distributions are $\mathbb{P}_Z$ and $\mathbb{P}_{X+\varepsilon}$. It is known that (66) is a metric for convergence in distribution among distributions with finite second moments, which means that it satisfies the axioms of distance and that $W_2(\nu_k, \mu) \to 0$ if and only if $\nu_k \xrightarrow{d} \mu$. Then, (65) implies that

$$W_2(\mathbb{P}_Z, \mathbb{P}_{X+\varepsilon}) = 0.$$

Based on this result, Arellano and Bonhomme (2021) obtain pseudo-observations of X by minimizing the sample version of (66).

I apply their approach to the panel data setting to obtain pseudo-observations of $\tilde{Y}_{it}$. I assume that ε_{it} follows an i.i.d. zero-mean Gaussian distribution such that $\text{Var}(\varepsilon_{it}) = 0.047$, which is the variance estimate of the transitory income process in Guvenen (2009).

[7] They also consider a more general case of multivariate factor models.

I then simulate $K = 200$ i.i.d. draws of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$:

$$\varepsilon_k = (\varepsilon_{k1}, \ldots, \varepsilon_{kT}) \in \mathbb{R}^T, \quad k = 1, \ldots, K.$$

Then, given the initial values of $\tilde{Y}_{it}$, denoted by

$$\tilde{Y}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iT}) \in \mathbb{R}^T, \quad i = 1, \ldots, N,$$

I obtain the *synthetic* data of Y_{it} by calculating:

$$\hat{Y}_{ik} = \tilde{Y}_i + \varepsilon_k \in \mathbb{R}^T, \quad i = 1, \ldots, N, \quad k = 1, \ldots, K, \tag{67}$$

where the synthetic data $\hat{Y}_{it}$ has size NK. Note that (67) computes a convolution of $\tilde{Y}_i$ and ε_{it}, because the distribution of $\hat{Y}_{ik}$ is equal to the convolution of the empirical distribution of $\tilde{Y}_i$ and the empirical distribution of ε_k.

I then compare the distribution of $\hat{Y}_{ik}$ with the observed distribution of Y_{it}. Let $\hat{P}_{\hat{Y}}$ and $\hat{P}_Y$ be the empirical distributions of $\hat{Y}_{ik}$ and Y_i, respectively. Then the (squared) second-order Wasserstein distance between the synthetic and the observed data is given by:

$$\begin{aligned} W_2^2(\hat{P}_Y, \hat{P}_{\hat{Y}}) = \min_{0 \leq p_{ijk} \leq 1} & \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{K} p_{ijk} ||Y_i - \hat{Y}_{jk}||^2 \\ \text{subject to} \qquad & \sum_{i=1}^{N} p_{ijk} = 1, \quad \sum_{j=1}^{N} \sum_{k=1}^{K} p_{ijk} = 1, \end{aligned}$$

which is the sample analogue of (66) and takes the form of a linear program. I then obtain pseudo-observations of $\tilde{Y}_{it}$, denoted by $\tilde{Y}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iT})$ for $i = 1, \ldots, N$, by:

$$\{\tilde{Y}_i\}_{i=1}^N = \operatorname*{argmin}_{\tilde{Y}_1, \ldots, \tilde{Y}_N} W_2^2(\hat{P}_Y, \hat{P}_{\hat{Y}}),$$

which can be shown to be a convex optimization problem. The minimizer of this problem is then the pseudo-observations $\tilde{Y}_i$.