Estimation with Aggregate Shocks

Jinyong Hahn* Guido Kuersteiner† Maurizio Mazzocco‡

UCLA University of Maryland UCLA

August 27, 2018

Abstract

Aggregate shocks affect most households' and firms' decisions. Using three stylized models we show that inference based on cross-sectional data alone generally fails to correctly account for decision making of rational agents facing aggregate uncertainty. We propose an econometric framework that overcomes these problems by explicitly parameterizing the agents' decision problem relative to aggregate shocks. Our framework and examples illustrate that the cross-sectional and time-series aspects of the model are often interdependent. Therefore, estimation of model parameters in the presence of aggregate shocks requires the combined use of cross-sectional and time-series data. We provide easy-to-use formulas for test statistics and confidence intervals that account for the interaction between the cross-sectional and time-series variation. Lastly, we perform Monte Carlo simulations that highlight the properties of the proposed method and the risks of not properly accounting for the presence of aggregate shocks.

*UCLA, Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095, hahn@econ.ucla.edu

†University of Maryland, Department of Economics, Tydings Hall 3145, College Park, MD, 20742, kuersteiner@econ.umd.edu

‡UCLA, Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095, mmazzocc@econ.ucla.edu

1 Introduction

An extensive body of economic research suggests that aggregate shocks have important effects on households' and firms' decisions. Consider for instance the oil shocks that hit developed countries in the seventies. Economists have argued that these aggregate shocks were responsible for recessions, periods of high inflation, low productivity, and reduced economic growth (Barsky and Lutz (2004)).

The profession has generally adopted one of the following three strategies to deal with aggregate shocks. The most common strategy is to assume that aggregate shocks have no effect on households' and firms' decisions, and hence that aggregate shocks can be ignored. Almost all papers estimating discrete choice dynamic models or dynamic games are based on this premise. Examples include Keane and Wolpin (1997), Bajari, Bankard, and Levin (2007), and Eckstein and Lifshitz (2011). The second approach is to add time dummies to the model in an attempt to capture the effect of aggregate shocks on the estimation of the parameters of interest, as was done for instance in Altug and Miller (1990), Runkle, (1991) and Shea (1995). The third strategy is to fully specify how aggregate shocks affect individual decisions within the structure of the economic problem. We are aware of only two papers that use this strategy, Lee and Wolpin (2010) and Dix-Carneiro (2014).

The previous discussion reveals that there is no generally agreed upon econometric framework for estimation and statistical inference in models where aggregate shocks affect individual decisions. This paper makes two main contributions related to this deficiency. We first develop three examples and use them to evaluate the effect of ignoring aggregate shocks on parameter estimation and the corresponding statistical inference. The examples provide important insights on what econometric approaches can be employed for the estimation of model parameters when aggregate shocks are present. Using these insights, we propose a method based on combining cross-sectional data with a long time series of aggregate variables. There are no available formulas for standard errors that can be used for statistical inference when these two data sources are combined. The second contribution is to provide simple-to-use test statistics and confidence intervals that are valid when our proposed method of combining cross-sectional and time-series data is used.

We proceed in four steps. In Section 2, we present a general class of models for which the presence of aggregate shocks generates the estimation and inferential issues studied in this paper. The models have two main features. First, each model in this class can be decomposed into two submodels. One submodel consists of all variables and parameters that can be studied using cross-sectional data. The other submodel includes all the variables and parameters that can be examined using time-series data. Second, each model in our class is based on decision making that depends on aggregate shocks and the parameters that govern their law of motion.

In our class of models the interactions between the two submodels and, hence, the effects of aggregate shocks on parameter estimation and inference can be complicated. To better understand these effects, in the second step, we present three examples that illustrate the complexities generated by the presence of aggregate shocks.

In Section 3, we consider as a first example a simple model of portfolio choice with aggregate shocks. The simplicity of the model enables us to clearly illustrate the effect of aggregates shocks on the estimation of model parameters and on their asymptotic distribution. Using the example, we first show that, if the econometrician does not properly account for the uncertainty generated by aggregate shocks, the estimates of model parameters are inconsistent. This result is noteworthy because, if the econometrician uses only cross-sectional data, it holds even if the model is correctly specified. The main implication of this result is that the inclusion of time dummies generally does not solve the issues introduced by the presence of aggregate shocks.[1] We then show that a method based on a combination of cross-sectional and time-series variables produces consistent estimates of the model parameters.

In Section 4, as a second example, we study the estimation of firms' production functions when aggregate shocks affect firms' decisions. This example shows that there are special cases where model parameters can be consistently estimated using only repeated cross-sections if time dummies are skillfully used rather than simply added as time intercepts. Specifically, our analysis indicates that the method proposed by Olley and Pakes (1996) fails to produce consistent estimates if aggregate shocks are present. It also indicates that the production functions can be consistently estimated if their method is modified with the proper inclusion of time dummies. The results of

[1] In the Euler equation context, Chamberlain (1984) considers a special example characterized by a nonstationary aggregate environment and time-varying nonstochastic preference shocks. Under this special environment, he shows that, when aggregate shocks are present but disregarded, the estimated parameters can be inconsistent even when time dummies are included. In this paper, we show that the presence of aggregate shocks produces such problems even in more general and realistic contexts.

Section 4 are of independent interest since aggregate shocks have significant effects in most markets and the estimation of firms' production functions is an important topic in industrial organization (Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015)).

In Section 5, we present as our last example a general equilibrium model of education and labor supply decisions. The portfolio example has the quality of being simple. But, because of its simplicity, it generates a unidirectional relationship between the time-series and cross-sectional submodels: the parameters of the cross-sectional submodel can be consistently estimated only if the parameters of the time-series submodel are known, but the time-series parameters can be consistently estimated without knowledge of cross-sectional parameters. However, this is not generally the case. In many situations, such as in Lee and Wolpin's (2010) paper, the link between the two submodels is bi-directional. The general-equilibrium example illustrates how a bi-directional relationship can arise. We use it for two purposes: to document, in a general setting, the complexity of the effects that aggregate shocks can have on estimation of model parameters and on their asymptotic distribution; and to explain how our method based on cross-sectional and time-series variables can be extended when the link between the two sub-models is bi-directional.

The examples make clear that, in general, consistent estimation of parameters in models with aggregate shocks requires the combined use of cross-sectional and time series data. There are no existing formulas for standard errors when these two data sources are combined. As the third step, in Section 6 we provide easy-to-use algorithms for test statistics and confidence intervals when parameter estimates are based on combined cross-sectional and time series data. The underlying asymptotic theory, which is presented in our companion paper Hahn, Kuersteiner, and Mazzocco (2016), is highly technical due to the complicated interactions that exists between the two submodels. Yet the resulting test statistics and confidence intervals take simple forms that are easy to apply. We conclude the section by discussing, using the portfolio choice model and the general equilibrium model, how standard errors can be computed in specific cases.

Finally, to evaluate our econometric framework, we perform a Monte Carlo experiment for the general equilibrium model. The Monte Carlo results indicate that our method performs well when the length of the time-series is sufficiently large. In that case, the parameter estimates are statistically close to the true values and the coverage probabilities are statistically close to the nominal levels. To document the effect of using only cross-sectional variation, we also compute coverage probabilities using the correct model, but under the assumption that the parameters that govern the law of motion of the aggregate shocks are known and not estimated. In this scenario, the coverage probabilities are computed without taking into account the variation of the time series estimates. We find that the difference between the true and erroneous coverage probabilities is generally large. Lastly, to evaluate the effect of ignoring aggregate shocks, we estimate the model's parameters under the incorrect assumption that the economy is not affected by aggregate shocks using cross-sectional data alone. Our results show that this form of misspecification can generate extremely large biases for the parameters that require both cross-sectional and longitudinal variation to be consistently estimated. For instance, we find that a parameter that is of considerable interest to economists, the coefficient of risk aversion, is between five and six times larger than the true value if aggregate shocks are ignored.

In addition to the econometric literature that deals with inferential issues, our paper also contributes to a growing literature whose objective is to estimate general equilibrium models. Some examples are Heckman and Sedlacek (1985), Heckman, Lochner, and Taber (1998), Lee (2005), Lee and Wolpin (2006), Gemici and Wiswall (2011), and Gillingham, Iskhakov, Munk-Nielsen, Rust, and Schjerning (2015). Aggregate shocks are an important feature of most general equilibrium models. Without those shocks, these models have the unpleasant implication that all aggregate variables can be fully explained by observables and, hence, that errors have no effect on those aggregate variables. Our general econometric framework makes this point clear by highlighting the impact of aggregate shocks on parameter estimation and the variation required in the data to estimate those models. More importantly, our results provide easy-to-use formulas that can be employed to perform statistical inference in a general equilibrium context.

A separate discussion is required for the papers by Lee and Wolpin (2006) and Dix-Carneiro (2014). These are the only papers we are aware of that estimate a model that fully specifies the effects of aggregate shocks on individual decision making. This allows the authors to obtain consistent estimates of the parameters of interest. These two papers are primarily focused on the estimation of a specific empirical model. They do not address the broader question of which statistical procedures and what type of data are needed to obtain consistent estimators when aggregate shocks are present. The focus of this paper is to answer these questions in the context of a general class of models.

2 A General Class of Models

This section presents a class of models for which the presence of aggregate shocks introduces the estimation and inferential issues that can be addressed by the approach we develop. Each model in the class has two main features. First, the model can be divided into two parts. One part encompasses all the variables and parameters of the model that can be analyzed using cross-sectional data and will be denoted with the term cross-sectional submodel. The other part includes the variables and parameters whose examination requires time-series data and will be denoted with the term time-series submodel. Second, the parameters of our models, which include the law of motion of the aggregate shocks, can be consistently estimated only if a combination of cross-sectional and time-series data are available.

Formally, each model consists of two distinct vectors of variables y_i and z_s. The first vector y_i includes all the variables that characterize the cross-sectional submodel. We allow for the possibility that the variables in y_i are collected from a short panel, in which case the vector y_i consists of several $y_{i,t}$'s, where the subscript t denotes the period in which $y_{i,t}$ is observed. Even when y_i is collected from a single cross section and hence $y_i = y_{i,t}$, it is sometimes useful to explicitly report the time period in which y_i is collected and write the cross-sectional vector as $y_{i,t}$. An example of a variable that belongs to $y_{i,t}$ is provided in the portfolio model that we present in Section 3. There, the share of resources invested in the risk-free asset by household i in period t is included in $y_{i,t}$. The second vector z_s is composed of all the variables associated with the time-series submodel. In the portfolio model of Section 3, the only variable in z_s is the aggregate return on the risky asset in period s.

We use two distinct time indices for the cross-sections (t) and time-series (s), because the periods in which the short panel is observed may differ from the periods in which the time series is observed. Specifically, we assume that our cross-sectional data consist of $\{y_{i,t},\ i = 1, \ldots, n,$ $t = T_0 + 1, \ldots, T_0 + T\}$, and our time-series data consist of $\{z_s,\ s = \tau_0 + 1, \ldots, \tau_0 + \tau\}$, where the interval $T_0 + 1, \ldots, T_0 + T$ may differ from the interval $\tau_0 + 1, \ldots, \tau_0 + \tau$. Situations in which the cross-sections and time-series cover distinct time intervals occur frequently. For example, the share of household resources invested in a risk-free asset, which is part of $y_{i,t}$ in the portfolio model, is observed in the Panel Studies of Income Dynamics (PSID) from 2001 to 2015 every two years.

But the aggregate return on the risky asset, measured as the return on the Standard & Poor's 500 index, which is included in z_s, is available every year from its inception in 1928. To simplify the notation, we adopt the normalization $T_0 = 0$.

The parameters of the general model can also be divided into two sets, β and ρ. The first set of parameters β characterizes the cross-sectional submodel, in the sense that, if the second set ρ was known, β and the vector of aggregate shocks ν_t could be consistently estimated using exclusively variation in the cross-sectional variables $y_{i,t}$. Similarly, the vector ρ characterizes the time-series submodel in the sense that, if β were known, the parameters ρ could be consistently estimated using exclusively the time series variables z_s. In many cases, ρ parametrizes the law of motion of the aggregate shocks.

There are two functions that relate the cross-sectional and time-series variables to the parameters. The function $f\left(y_{i,t}\middle|\,\beta,\nu,\rho\right)$ restricts the behavior of the cross-sectional variables conditional on a particular value of the parameters. Analogously, the function $g\left(z_s\middle|\,\beta,\rho\right)$ describes the behavior of the time-series variables for a given value of the parameters. The portfolio model discussed in Section 3 provides examples of such functions. In that model, (i) the variables $y_{i,t}$ are i.i.d. given the aggregate shock ν_t, (ii) the variables z_s correspond to (ν_s, ν_{s-1}), where ν_s denotes the realization of the aggregate shock in the s-th period, (iii) the cross-sectional function $f\left(y_{i,t}\middle|\,\beta,\nu_t,\rho\right)$ denotes the log likelihood of $y_{i,t}$ given the aggregate shock ν_t, and (iv) the time-series function $g\left(z_s\middle|\,\beta,\rho\right) = g\left(\nu_s\middle|\,\nu_{s-1},\rho\right)$ is the log of the conditional probability density function of the aggregate shock ν_s given ν_{s-1}. In this special case, the time-series function g does not depend on the cross-sectional parameters β.

The parameters of the general model can be estimated by maximizing a well-specified objective function. Since our framework consists of two submodels, a natural approach is to estimate the parameters of interest by maximizing two separate objective functions, one for the cross-sectional submodel and one for the time-series submodel. We denote these criterion functions by $F_n(\beta,\nu,\rho)$ and $G_\tau(\beta,\rho)$. In the case of maximum likelihood, the criterion functions are simply $F_n(\beta,\nu,\rho) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} f\left(y_{i,t}\middle|\,\beta,\nu_t,\rho\right)$ and $G_\tau(\beta,\rho) = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} g\left(z_s\middle|\,\beta,\rho\right)$,where $\nu = (\nu_1, ..., \nu_T)$. Another scenario where separate criterion functions arise naturally is when f and g represent moment conditions. The use of two separate objective functions is helpful in our context because it enables us to discuss which issues arise if only cross-sectional or only time-series variables are used in the estimation. More importantly from a practical viewpoint, considering the two components separately adds flexibility to the parameter estimation since, in some periods, data may be required to construct the variables of only one of the submodels.

In our class of models, the cross-sectional submodel suffers from a source of identification failure that has received little or no attention in the literature: because ν_t does not vary in the cross-section, $F_n(\beta, \nu, \rho)$ does not separately identify ρ and converges to $F(\beta, \nu, \rho) = T^{-1} \sum_{t=1}^{T} E\left[f\left(y_{i,t} \middle| \beta, \nu_t, \rho\right) | \nu_{0,t}\right]$, where $\nu_{0,t}$ is used to denote the true realization of the shock. Namely, the limit $F(\beta, \nu, \rho)$ of the criterion function remains random, as it depends on the random variables $\nu_{0,t}$. The main implication is that, generally, estimators obtained from maximizing $F(\beta, \nu, \rho)$ are functions of the aggregate shock and, therefore, are not consistent since they vary with these shocks. Because of this, the class of models we study in this paper requires both cross-sectional and time-series data for the identification of the parameters.

The portfolio model of Section 3 and the general equilibrium model of Section 5 represent examples where this identification problem arises. They consider two different situations. In the portfolio model, the relationship between the cross-sectional and time-series submodels is unidirectional since, as we indicate above, the time-series function g does not depend on the cross-sectional parameters β, but the cross-sectional function f does depend on the time-series parameters ρ. Moreover, ρ cannot be identified using cross-sectional data alone, because the maximum of the cross-sectional objective function F over the parameters β and the aggregate variables ν does not change with the value of ρ. However, ρ can be identified using time-series data. Therefore, the cross-sectional submodel suffers from the identification problem, as β can be consistently estimated only if time-series data are first used to identify ρ. In this case, we propose an empirical strategy based on a two-step procedure that first identifies ρ from time series data and then uses the estimated ρ in combination with cross-sectional data to identify β. The general equilibrium model describes a situation in which the two-step procedure fails, as the time-series function g depends on the cross-sectional parameters β and the cross-sectional function f depends on the time-series parameters ρ. In this situation, we propose a procedure that simultaneously identifies β and ρ, by combining cross-sectional and time-series data.

3 Example 1: Portfolio Choice

We start with a simple portfolio choice example that clearly illustrates the perils of ignoring aggregate shocks. Using this example, we make the following points. First, the presence of aggregate shocks generally produces inconsistent parameter estimates unless the econometrician properly accounts for the uncertainty generated by the aggregate shocks. In fact, many parameters cannot be identified using cross-sectional data alone, even if the model is correctly specified. Second, the use of time dummies generally does not solve the problems generated by the existence of aggregate shocks. Third, if the researcher does not account for the aggregate shocks, the parameter estimates will adjust to make the model consistent with the aggregate uncertainty that is present in the data but not modeled, hence the inconsistency. For instance, in a model with risk averse agents such as our portfolio example, ignoring the aggregate shocks produces estimates of the risk aversion parameter that are upward biased.

Consider an economy that, in each period t, is populated by n households. These households are born at the beginning of period t, live for one period, and are replaced in the next period by n new families. The households living in consecutive periods do not overlap and, hence, make independent decisions. Each household is endowed with deterministic income and has preferences over a non-durable consumption good $c_{i,t}$. The preferences can be represented by Constant Absolute Risk Aversion (CARA) utility functions which take the following form: $U\left(c_{i,t}\right) = -e^{-\delta c_{i,t}}$. For simplicity, we normalize income to be equal to 1.

During the period in which households are alive, they can invest a share of their income in a risky asset with return $u_{i,t}$. The remaining share is automatically invested in a risk-free asset with a return r that does not change over time. At the end of the period, the return on the investment is realized and households consume the quantity of the non-durable good they can purchase with their realized income. The return on the risky asset depends on aggregate shocks. Specifically, it takes the following form: $u_{i,t} = \nu_t + \epsilon_{i,t}$, where ν_t is the aggregate shock and $\epsilon_{i,t}$ is an i.i.d. idiosyncratic shock. The idiosyncratic shock, and hence the heterogeneity in the return on the risky asset, can be interpreted as differences across households in transaction costs, in information on the profitability of different stocks, or in marginal tax rates. We assume that $\nu_t \sim N\left(\mu, \sigma_\nu^2\right)$, $\epsilon_{i,t} \sim N\left(0, \sigma_\epsilon^2\right)$, and hence that $u_{i,t} \sim N\left(\mu, \sigma^2\right)$, where $\sigma^2 = \sigma_\nu^2 + \sigma_\epsilon^2$.

Household i living in period t chooses the fraction of income to be allocated to the risk-free asset $\alpha_{i,t}$ by maximizing its life-time expected utility[2]:

$$\max_{\alpha_{i,t}} \; E\left[-e^{-\delta c_{i,t}}\right]$$
$$s.t. \quad c_{i,t} = \alpha_{i,t}\left(1+r\right) + \left(1-\alpha_{i,t}\right)\left(1+u_{i,t}\right), \tag{1}$$

where the expectation is taken with respect to the return on the risky asset. It can be shown[3] that the household's optimal choice of $\alpha_{i,t}$ is given by

$$\alpha^*_{i,t} = \alpha = \frac{\delta\sigma^2 + r - \mu}{\delta\sigma^2}. \tag{2}$$

We assume that the econometrician is mainly interested in estimating the risk aversion parameter δ.

We now consider an estimator that takes the form of a sample analog of (2), and study the impact of aggregate shocks on the estimator's consistency when an econometrician ignores the aggregate shocks and works only with cross-sectional data. Our analysis reveals that such an estimator is inconsistent because cross-sectional data do not contain information about aggregate uncertainty. It also makes explicit the dependence of the estimator on the probability distribution of the aggregate shock and thus points to the following method for consistently estimating δ. First, using time series variation, the parameters pertaining to aggregate uncertainty are consistently estimated. Second, those estimates are plugged into the cross-sectional model to estimate the remaining parameters.[4]

Without loss of generality, we assume that the cross-sectional data are observed in period $t = 1$. The econometrician observes data on the return of the risky asset $u_{i,t}$ and on the return of

[2]Our models assume rational expectations. We do not consider examples that incorporate model uncertainty, i.e., the possibility that agents need to learn or estimate model parameters when making decisions. We restrict our attention to rational expectation models because there is only a limited number of papers that consider self-confirming equilibria or robust control. See Cho, Sargent, and Williams (2002) or Hansen, Sargent, and Tallarini (1999). This is, however, an important topic that we leave for future research.

[3]This is shown in the on-line Appendix.

[4]Our model is a stylized version of many models considered in a large literature interested in estimating the parameter δ using cross-sectional variation. Estimators are often based on moment conditions derived from first order conditions (FOC) related to optimal investment and consumption decisions. Such estimators have similar problems, which we discuss in the on-line Appendix A.2.

the risk-free asset r. We assume that they also observe a noisy measure of the share of resources invested in the risk-free asset $\alpha_{i,t} = \alpha + e_{i,t}$, where $e_{i,t}$ is a measurement error with zero mean and variance σ_e^2. The vector of cross-sectional variables y_i is therefore composed of u_{i1} and α_{i1} and the vector of cross-sectional parameters β is composed of δ, σ_ϵ^2, and σ_e^2. The vector of time-series variables includes only the aggregate shock, i.e. $z_t = \nu_t$, and the vector of time-series variables parameters is composed of μ and σ_ν^2. Since, ν_t corresponds to the aggregate return of the risky asset, we assume that ν_t is observed.

Suppose that the econometrician ignores the existence of the aggregate shocks, by assuming that the aggregate return is fixed at μ for all t, and uses only cross-sectional variation. Recall that $\mu = E[u_{i1}]$, $\sigma^2 = \mathrm{Var}(u_{i1})$, and $\alpha = E[\alpha_{i1}]$. The econometrician will therefore estimate those parameters using the following method-of-moments estimators:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} u_{i1} = \bar{u}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (u_{i1} - \bar{u})^2, \qquad \text{and} \qquad \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{i1}.$$

The econometrician can then use equation (2) to write the risk aversion parameter as $\delta = (\mu - r)/(\sigma^2(1-\alpha))$ and estimate it with the sample analog $\hat{\delta} = (\hat{\mu} - r)/(\hat{\sigma}^2(1-\hat{\alpha}))$.

In the presence of the aggregate shocks ν_t, however, the method-of-moments estimators take the following form:

$$\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} u_{i1} = \nu_1 + \frac{1}{n}\sum_{i=1}^{n} \epsilon_{i1} = \nu_1 + o_p(1), \\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} (u_{i1} - \bar{u})^2 = \frac{1}{n}\sum_{i=1}^{n} (\epsilon_{i1} - \bar{\epsilon})^2 = \sigma_\epsilon^2 + o_p(1), \\
\hat{\alpha} &= \alpha + \frac{1}{n}\sum_{i=1}^{n} e_{i1} = \alpha + o_p(1),
\end{aligned}$$

which implies that δ will be estimated to be

$$\hat{\delta} = \frac{\nu_1 + o_p(1) - r}{(\sigma_\epsilon^2 + o_p(1))(1 - \alpha + o_p(1))} = \frac{\nu_1 - r}{\sigma_\epsilon^2(1-\alpha)} + o_p(1). \tag{3}$$

Using equation (3), we can study the properties of estimator $\hat{\delta}$. Without aggregate shocks, we would have $\nu_1 = \mu$, $\sigma_\nu^2 = 0$, $\sigma_\epsilon^2 = \sigma^2$ and, therefore, $\hat{\delta}$ would converge to δ, a nonstochastic constant, as n grows to infinity. It is therefore a consistent estimator of the risk aversion parameter. However, in the presence of the aggregate shock, the proposed estimator has different properties. We consider first the case in which econometricians condition on the realization of the aggregate shock ν or, equivalently, assume that the realization of the aggregate shock is known. In this case, the estimator $\hat{\delta}$ is inconsistent with probability 1, since it converges to $\frac{\nu_1 - r}{\sigma_\epsilon^2(1-\alpha)}$ and not to the true value $\frac{\mu - r}{(\sigma_\nu^2+\sigma_\epsilon^2)(1-\alpha)}$.

As discussed in the introduction, a common practice to account for the effect of aggregate shocks is to include time dummies in the model. The portfolio example clarifies that the addition of time dummies does not solve the problem generated by the presence of aggregate shocks. The inclusion of time dummies is equivalent to the assumption that the realization of the aggregate shock is known or that econometricians condition on the realization of ν. But the previous result indicates that, using exclusively cross-sectional data, the estimator $\hat{\delta}$ is inconsistent even if the realizations of the aggregate shocks are known. To provide the intuition behind this result, note that, if aggregate shocks affect individual behavior, the decisions recorded in the data account for the uncertainty generated by the variation in ν. Even if econometricians assume that the realizations of the aggregate shocks are known, the only way the portfolio model can rationalize the degree of uncertainty displayed by the data is by making the agents more risk averse than they actually are. Hence, the inconsistency described above.

We now consider the case in which econometricians do not condition on the realization of the aggregate shock. As n grows to infinity, $\hat{\delta}$ converges to a random variable with a mean that is different from the true value of the risk aversion parameter. The estimator will therefore be inconsistent. To see this, remember that $\nu_1 \sim N(\mu, \sigma_\nu^2)$. As a consequence, the unconditional asymptotic distribution of $\hat{\delta}$ takes the following form:

$$\hat{\delta} \to_d N\left(\frac{\mu - r}{\sigma_\epsilon^2(1-\alpha)}, \left(\frac{1}{\sigma_\epsilon^2(1-\alpha)}\right)^2 \sigma_\nu^2\right) = N\left(\delta + \delta\frac{\sigma_\nu^2}{\sigma_\epsilon^2}, \frac{\sigma_\nu^2}{(\sigma_\epsilon^2(\alpha-1))^2}\right),$$

which is centered at $\delta + \delta\frac{\sigma_\nu^2}{\sigma_\epsilon^2}$ and not at δ, hence the asymptotic bias. The intuition behind the asymptotic bias is the same as for the case in which the realization of the aggregate shock is known. But when econometricians do not condition on ν, it is straightforward to sign the asymptotic bias. The asymptotic bias is equal to $\delta \frac{\sigma_\nu^2}{\sigma_\epsilon^2}$ and always positive, which is consistent with the intuition described above according to which ignoring aggregate shocks generates estimates of the risk aversion parameter that are too high. The formula of the asymptotic bias also enables one to reach the intuitive conclusion that its size increases when the magnitude of the aggregate uncertainty (σ_ν^2) is large relative to the magnitude of the micro-level uncertainty (σ_ϵ^2).[5]

In our example, the statistical uncertainty, captured by the variance of the estimator, does not vanish with the sample size and, hence, $\hat{\delta}$ converges to a random variable. We are not the first to consider a case in which common factors affect the limiting distribution of an estimator. Andrews (2005) and more recently Kuersteiner and Prucha (2013) discuss similar scenarios. However, in their case the common factor enters in a more restrictive way essentially only affecting the variance but not the mean of a regression error. In our example on the other hand aggregate variables have a profound effect on model specification. As a result the nature of the asymptotic randomness is such that the estimator is not even consistent or asymptotically unbiased. This is not the case in Andrews (2005) or Kuersteiner and Prucha (2013), where the asymptotic randomness affects the variance of the limiting distribution of the estimator but not the asymptotic bias or its consistency.[6]

As mentioned above, there is a simple statistical explanation for our result: cross-sectional variation is not sufficient for the consistent estimation of the risk aversion parameter if aggregate shocks affect individual decisions. To make this point transparent, observe that, conditional on the aggregate shock, the assumptions of this section imply that the cross-sectional variable $y_i = (u_{i1}, \alpha_{i1})$ has the following distribution:

$$y_i | \, \nu_1 \sim N \left(\left[\begin{array}{c} \nu_1 \\ \dfrac{\delta\left(\sigma_\nu^2 + \sigma_\epsilon^2\right) + r - \mu}{\delta\left(\sigma_\nu^2 + \sigma_\epsilon^2\right)} \end{array} \right], \left[\begin{array}{cc} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_e^2 \end{array} \right] \right). \tag{4}$$

Using (4), it is straightforward to see that any arbitrary choice of the time-series parameters $\rho = (\mu, \sigma_\nu^2)$ maximize the cross-sectional likelihood, as long as one chooses δ that satisfies the

[5] When the realization of ν is assumed to be known, one can only sign the expected bias, where the expectation is taken over the realization of the aggregate shock, since the bias depends on the actual realization of the shock. The expected bias is always positive and increasing in σ_ν^2 as our intuition indicates.

[6] Kuersteiner and Prucha (2013) also consider cases where the estimator is random and inconsistent. However, in their case this happens for different reasons: the endogeneity of the factors. The inconsistency considered here occurs even when the factors are strictly exogenous.

following equation:

$$\frac{\delta\left(\sigma_{\nu}^{2}+\sigma_{\epsilon}^{2}\right)+r-\mu}{\delta\left(\sigma_{\nu}^{2}+\sigma_{\epsilon}^{2}\right)}=\alpha.$$

Consequently, the cross-sectional parameters μ and σ_{ν}^{2} cannot be consistently estimated by maximizing the cross-sectional likelihood and, hence, δ cannot be consistently estimated using only cross-sectional data.

We can now describe the method we propose in this paper as a general solution to the issues introduced by the presence of aggregate shocks. The method, which generates consistent estimates of the model parameters, relies on the combined use of cross-sectional and time-series variables. Specifically, under the assumption that the realizations of the aggregate shocks are observed, the researcher can consistently estimate the parameters that characterize the distribution of those shocks μ and σ_{ν}^{2} using a time-series of aggregate data $\{z_s\}$.[7] The risk aversion parameter δ and the remaining two parameters σ_{ϵ}^{2} and σ_{e}^{2} can then be consistently estimated using cross-sectional variables, by replacing the consistent estimators of μ and σ_{ν}^{2} in the correctly specified cross-sectional likelihood derived in equation (4).

The application of the method we propose to the estimation of the portfolio model can also be illustrated using the general notation introduced in Section 2. The cross-sectional function $f\left(\cdot\middle|\,\beta,\nu_t,\rho\right)$ corresponds to the log of the density of the variable $y_{i,t}$ defined in (4), and the time-series function $g\left(\cdot\middle|\,\beta,\rho\right)$ corresponds to the conditional density of the aggregate shock v_s given v_{s-1}, where $\beta=(\delta,\sigma_{\epsilon}^{2},\sigma_{e}^{2})$ and $\rho=(\mu,\sigma_{\nu}^{2})$.

The general model of Section 2 applied to the portfolio example can also be used to illustrate how the use of cross-sectional data alone generates an asymptotic bias, whether the model is correctly or incorrectly specified. Denote by ϕ the standard normal density. Then the limit of the correctly specified cross-sectional log likelihood is $E\left[\log\left(\phi\left(\left(y_{i1}-\mu\right)/\sigma\right)/\sigma\right)|\nu_1\right]+E\left[\log\left(\phi\left(\left(\alpha_{i1}-\alpha\right)/\sigma_e\right)/\sigma_e\right)\right]$, which is maximized at the inconsistent values $\mu=\nu_1$ and $\sigma=\sigma_{\epsilon}$. Now imagine that the fictitious econometrician, relying solely on cross-sectional variation, misspecifies the function $f\left(\cdot\middle|\,\beta,\nu_t,\rho\right)$ by setting $\mu=\nu_1$ and $\sigma_{\nu}^{2}=0$. In the context of the model considered in this section, this leads to the same inconsistent cross-sectional estimator as under

[7]The assumption that the realizations of aggregate shocks are observed is made to simplify the discussion and can be easily relaxed. In Section 5, we apply the proposed estimation method to a general equilibrium example in which the realizations of the aggregate shocks are not observed.

the correct specification.

The example presented in this section is a simplified version of the general class of models introduced in Section 2. The variables and parameters of the time-series model affect the cross-sectional model, but the cross-sectional variables and parameters have no impact on the time-series model. As a consequence, the time-series parameters can be consistently estimated without knowing the cross-sectional parameters. The recursive feature of the example is due to the exogenously specified price process and the partial equilibrium nature of the model. In more complicated situations, such as general equilibrium models, where aggregate shocks are a natural feature, the relationship between the two models is generally bi-directional. But before considering an example of the general case, we study a situation in which the effect of aggregate shocks can be accounted for with the proper use of time dummies.

4 Example 2: Estimation of Production Functions

In the previous section, we presented an example that illustrates the complicated nature of identification in the presence of aggregate shocks. The example highlights that generally there is no simple method for estimating the class of models considered in this paper. Estimation requires a careful examination of the interplay between the cross-sectional and time-series models. In this section, we consider an example showing that there are exceptions to this general rule. In the case we analyze, the researcher is interested in only a subset of the parameters, and its identification can be achieved using only cross-sectional data even if aggregate shocks affect individual decisions, provided that time dummies are skillfully employed. We show that the naive practice of introducing additive time dummies is not sufficient to deal with the effects generated by aggregate shocks. But the solution is simpler than the general approach we adopted to identify the parameters of the portfolio model.[8]

The example we consider here is a simplified version of the problem studied by Olley and Pakes (1996) and deals with an important topic in industrial organization: the estimation of firms' production functions. A profit-maximizing firm j produces a product $Y_{j,t}$ in period t, employing a

[8]Altug and Miller (1990) represents another example where the skillfull use of time dummies, economic assumptions, and functional form assumptions enables the researcher to consistently estimate the model parameters using cross-sectional data alone, even in the presence of aggregate shocks.

production function that depends on the logarithm of labor $l_{j,t}$, the logarithm of capital $k_{j,t}$, and a productivity shock $\omega_{j,t}$. By denoting the logarithm of $Y_{j,t}$ by $y_{j,t}$, the production function takes the following form:

$$y_{j,t} = \beta_0 + \beta_l l_{j,t} + \beta_k k_{j,t} + \omega_{j,t} + \eta_{j,t}, \tag{5}$$

where $\eta_{i,t}$ is a measurement error. The firm chooses the amount of labor to use in production and the new investment in capital $i_{j,t}$ by maximizing a dynamic profit function subject to the constraints that in each period capital accumulates according to the following equation:[9]

$$k_{j,t+1} = (1-\delta)\, k_{j,t} + i_{j,t},$$

where δ is the rate at which capital depreciates. In the model proposed by Olley and Pakes (1996), firms are heterogeneous in their age and can choose to exit the market. In this section, we abstract from age heterogeneity and exit decisions because they make the model more complicated without adding more insight on the effect of aggregate shocks on the estimation of production functions.

A crucial feature of the model proposed by Olley and Pakes (1996) and of our example is that the optimal investment decision in period t is a function of the current stock of capital and of the productivity shock, i.e.

$$i_{j,t} = i_t\left(\omega_{j,t}, k_{j,t}\right). \tag{6}$$

Olley and Pakes (1996) do not allow for aggregate shocks, but in this example we consider a situation in which the productivity shock at t is the sum of an aggregate shock ν_t drawn from a distribution $F\left(\nu \,|\rho\right)$ and of an idiosyncratic shock $\varepsilon_{j,t}$ independent of ν_t, i.e.

$$\omega_{j,t} = \nu_t + \varepsilon_{j,t}.$$

One example of aggregate shocks affecting the productivity of a firm is the arrival of technological innovations in the economy. We assume that ν_t and $\varepsilon_{j,t}$ are both Markov processes and that the firm observes the realization of the aggregate shock and, separately, of the idiosyncratic shock.

We first review the estimation method proposed by Olley and Pakes (1996) for the production function (5) when aggregate shocks are not present. We then discuss how that method can be

[9]For details of the profit function, see Olley and Pakes (1996).

modified with the appropriate use of time dummies if aggregate shocks affect firms' decisions.

The main problem in the estimation of the production function (5) is that the productivity shock is correlated with labor and capital, but not observed by the econometrician. To deal with that issue, Olley and Pakes (1996) use the result that the investment decision (6) is strictly increasing in the productivity shock for every value of capital to invert the corresponding function, solve for the productivity shock, and obtain

$$\omega_{j,t} = h_t\left(i_{j,t}, k_{j,t}\right). \tag{7}$$

One can then replace the productivity shock in the production function using equation (7) to obtain

$$y_{j,t} = \beta_l l_{j,t} + \phi_t\left(i_{j,t}, k_{j,t}\right) + \eta_{j,t}, \tag{8}$$

where

$$\phi_t\left(i_{j,t}, k_{j,t}\right) = \beta_0 + \beta_k k_{j,t} + h_t\left(i_{j,t}, k_{j,t}\right). \tag{9}$$

The parameter β_l and the function ϕ_t can then be estimated by regressing, period by period, $y_{j,t}$ on $l_{j,t}$ and a flexible polynomial (i.e., a nonparametric approximation) in $i_{j,t}$ and $k_{j,t}$ or, similarly, by interacting time dummies with the polynomial in $i_{j,t}$ and $k_{j,t}$.[10] The parameter β_l is therefore identified by[11]

$$\beta_l = \frac{E\left[\left(l_{j,t} - E\left[l_{j,t}\right| i_{j,t}, k_{j,t}]\right)\left(y_{j,t} - E\left[y_{j,t}\right| i_{j,t}, k_{j,t}]\right)\right]}{E\left[\left(l_{j,t} - E\left[l_{j,t}\right| i_{j,t}, k_{j,t}]\right)^2\right]}. \tag{10}$$

To identify the parameter on the logarithm of capital β_k observe that the production function

[10] Given our simplifying assumptions that there are no exit decisions and age heterogeneity, without aggregate shocks, the functions ϕ and h are independent of time. We use the more general notation that allows for time dependence to highlight that the estimation approach developed in Olley and Pakes (1996) generally fails when aggregate shocks are present, even if ϕ and h are allowed to depend on time.

[11] As noted by Ackerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2011), the Olley and Pakes's procedure is potentially subject to the functional dependence problem, i.e. without additional assumptions labor is a deterministic function of investment and capital and, hence, the denominator of (10) is equal to zero. To address this issue, we assume that $l_{j,t}$ is a function of some other exogenous variable in addition to $(i_{j,t}, k_{j,t})$. One possibility is to assume that the investment decision is made before the labor decision and that an unanticipated firm-specific shock to the price of labor is realized between the time of the investment and labor decisions. See Ackerberg, Caves, and Frazer (2015, pp. 2424-2427) for a detailed discussion.

(5) implies the following:

$$E\left[y_{i,t+1}-\beta_l l_{j,t+1}\middle|\, k_{j,t+1}\right]=\beta_0+\beta_k k_{j,t+1}+E\left[\omega_{j,t+1}\middle|\, \omega_{j,t}\right]=\beta_0+\beta_k k_{j,t+1}+g\left(\omega_{j,t}\right), \tag{11}$$

where the first equality follows from $k_{j,t+1}$ being determined conditional on $\omega_{j,t}$. Note that, in the absence of aggregate shocks, the function $g(\cdot)$ is independent of time. The shock $\omega_{j,t}=h_t\left(i_{j,t},k_{j,t}\right)$ is not observed, but using equations (7) and (9), it can be written in the following form:

$$\omega_{j,t}=\phi_t\left(i_{j,t},k_{j,t}\right)-\beta_0-\beta_k k_{j,t}, \tag{12}$$

where ϕ_t is known from the first-step estimation. Substituting for $\omega_{j,t}$ into the function $g(\cdot)$ in equation (11) and letting $\xi_{j,t+1}=\omega_{j,t+1}-E\left[\omega_{j,t+1}\middle|\, \omega_{j,t}\right]$, equation (11) can be written as follows:

$$y_{i,t+1}-\beta_l l_{j,t+1}=\beta_k k_{j,t+1}+g\left(\phi_t-\beta_k k_{j,t}\right)+\xi_{j,t+1}+\eta_{j,t}, \tag{13}$$

where β_0 has been included in the function $g(\cdot)$. The parameter β_k can then be estimated by using the estimates of β_l and ϕ_t obtained in the first step and by minimizing the sum of squared residuals in the previous equation, employing a kernel or a series estimator for the function g.

We now consider the case in which aggregate shocks affect the firm's decisions and analyze how the model parameters can be identified using only cross-sectional variation. The introduction of aggregate shocks changes the estimation method in two main ways. First, the investment decision is affected by the aggregate shock and takes the following form:

$$i_{j,t}=i_t\left(\nu_t,\varepsilon_{j,t},k_{j,t}\right).$$

Second, all expectations are conditional on the realization of the aggregate shock since in the cross-section there is no variation in that shock and only its realization is relevant.

The shocks ν_t and $\varepsilon_{j,t}$ enter as independent arguments in the investment function to maintain the assumption made in Olley and Pakes that the problem solved by the firm is Markovian. To understand why, consider a case in which ν_t and $\varepsilon_{j,t}$ are both AR(1) processes. If we only use their sum as a state variable, the Markovian assumption is generally violated, because the sum of

AR(1) processes is in general not an AR(1) but an ARMA(2,1) process. However, if we include ν_t and $\varepsilon_{j,t}$ as separate state variables – both observed by the firm – the Markovian structure is preserved.

If the investment function is strictly increasing in the productivity shock $\omega_{j,t}$ for all capital levels, it is also strictly increasing in ν_t and $\varepsilon_{j,t}$ for all $k_{j,t}$, because $\omega_{j,t} = \nu_t + \varepsilon_{j,t}$. Using this result, we can invert $i_t(\cdot)$ to derive $\varepsilon_{j,t}$ as a function of the aggregate shock, investment, and the stock of capital, i.e.

$$\varepsilon_{j,t} = h_t(\nu_t, i_{j,t}, k_{j,t}).$$

The production function can therefore be rewritten in the following form:

$$\begin{aligned} y_{j,t} &= \beta_0 + \beta_l l_{j,t} + \beta_k k_{j,t} + \nu_t + \varepsilon_{j,t} + \eta_{j,t} \\ &= \beta_l l_{j,t} + [\beta_0 + \beta_k k_{j,t} + \nu_t + h_t(\nu_t, i_{j,t}, k_{j,t})] + \eta_{j,t} \\ &= \beta_l l_{j,t} + \bar{\phi}_t(\nu_t, i_{j,t}, k_{j,t}) + \eta_{j,t} \\ &= \beta_l l_{j,t} + \phi_t(i_{j,t}, k_{j,t}) + \eta_{j,t}, \end{aligned} \tag{14}$$

where we have included the aggregate shock in the function ϕ_t. Analogously to the case of no aggregate shocks, β_l can be consistently estimated by regressing period by period $y_{j,t}$ on $l_{j,t}$ and a polynomial in $i_{j,t}$ and $k_{j,t}$ or, similarly, by interacting the polynomial with time dummies.

Note that estimation of β_l is not affected by uncertainty generated by the aggregate shocks since that uncertainty is captured by the time subscript in the function ϕ_t and the method developed by Olley and Pakes (1996) already requires the estimation of a different function ϕ for each period. The parameter β_l is therefore identified by[12]

$$\beta_l = \frac{E\left[\left(l_{j,t} - E\left[l_{j,t}\middle| i_{j,t}, k_{j,t}, \nu_t\right]\right)\left(y_{j,t} - E\left[y_{j,t}\middle| i_{j,t}, k_{j,t}, \nu_t\right]\right)|\nu_t\right]}{E\left[\left(l_{j,t} - E\left[l_{j,t}\middle| i_{j,t}, k_{j,t}, \nu_t\right]\right)^2 |\nu_t\right]}. \tag{15}$$

Observe that the expectation operator in the previous equation is in principle defined with respect to a probability distribution function that includes the randomness of the aggregate shock ν_t. But,

[12]We note that the expectations in (15) are conditional on ν_t. The reason is that ν_t is invariant in the cross-section. Convergence of cross-sectional averages to conditional expectations then is a consequence of the Ergodic Theorem.

when one uses cross-sectional variation, ν_t is fixed at its realized value. As a consequence, the distribution is only affected by the randomness of ε_{it}.

For the estimation of β_k, note that, under the assumption that the ν_t's are independent of the $\varepsilon_{j,t}$'s,

$$\begin{aligned} & E\left[y_{i,t+1} - \beta_l l_{j,t+1} \middle| k_{j,t+1}, i_{j,t}, k_{j,t}, \nu_{t+1}, \nu_t, \varepsilon_{j,t}\right] \\ & = \beta_0 + \beta_k k_{j,t+1} + E\left[\nu_{t+1} + \varepsilon_{j,t+1} \middle| k_{j,t}, \nu_{t+1}, \nu_t, \varepsilon_{j,t}\right] \\ & = \beta_0 + \beta_k k_{j,t+1} + \nu_{t+1} + E\left[\varepsilon_{j,t+1} \middle| \varepsilon_{j,t}\right] \\ & = \beta_0 + \beta_k k_{j,t+1} + \nu_{t+1} + g\left(\varepsilon_{j,t}\right), \end{aligned} \tag{16}$$

where the first equality follows from $k_{j,t+1}$ being known if $i_{j,t}$, $k_{j,t}$, ν_t, and $\varepsilon_{j,t}$ are known.

The only variable of equation (16) that is not observed is $\varepsilon_{j,t}$. But remember that

$$\varepsilon_{j,t} = h_t\left(\nu_t, i_{j,t}, k_{j,t}\right) = \phi_t\left(\nu_t, i_{j,t}, k_{j,t}\right) - \beta_0 - \beta_k k_{j,t} - \nu_t.$$

We can therefore use the above expression to substitute for $\varepsilon_{j,t}$ in equation (16) and obtain

$$\begin{aligned} & E\left[y_{i,t+1} - \beta_l l_{j,t+1} \middle| k_{j,t+1}, i_{j,t}, k_{j,t}, \nu_{t+1}, \nu_t\right] \\ & = \beta_0 + \beta_k k_{j,t+1} + \nu_{t+1} + g_t\left(\phi_t\left(\nu_t, i_{j,t}, k_{j,t}\right) - \beta_0 - \beta_k k_{j,t} - \nu_t\right) \\ & = \beta_k k_{j,t+1} + g_{t,t+1}\left(\phi_t - \beta_k k_{j,t}\right), \end{aligned}$$

where in the last equality β_0, ν_t, and ν_{t+1} have been included in the function $g_{t,t+1}(\cdot)$. Hence, if one defines $\xi_{j,t+1} = \varepsilon_{j,t+1} - E\left[\varepsilon_{j,t+1} \middle| \nu_t, \varepsilon_{j,t}\right]$, the previous equation can be written in the following form:

$$y_{i,t+1} - \beta_l l_{j,t+1} = \beta_k k_{j,t+1} + g_{t,t+1}\left(\phi_t - \beta_k k_{j,t}\right) + \xi_{j,t+1} + \eta_{j,t+1}. \tag{17}$$

The inclusion of the aggregate shocks in the function $g(\cdot)$ implies that that function varies with time when aggregate shocks are present. This is in contrast with the case considered in Olley and Pakes (1996) where aggregate shocks are ignored and, hence, the function $g(\cdot)$ is independent of time.

Given equation (17), if one attempts to estimate β_k using equation (13), repeated cross-sections and the method developed for the case with no aggregate shocks, the estimated coefficient will generally be inconsistent because the econometrician does not account for the aggregate shocks and their correlation with the firm's choice of capital. There is, however, a small variation of the method proposed earlier that produces consistent estimates of β_k, as long as $\varepsilon_{j,t}$ is independent of $\eta_{j,t}$. The econometrician should regress *period by period* $y_{j,t}$ on $l_{j,t}$ and a nonparametric function of $i_{j,t}$ and $k_{j,t}$ or, in practice, on a flexible polynomial of $i_{j,t}$ and $k_{j,t}$ interacted with time dummies. It is this atypical use of time dummies that enables the econometrician to account for the effect of aggregate shocks on firms' decisions.

We conclude by drawing attention to two features of the production function example that make it possible to use time dummies to deal with the effect of the aggregate shocks. To do that, it is useful to cast the example in terms of the cross-sectional and time-series models. The cross-sectional model includes the variables y_j, l_j, k_j, and i_j, the parameters β_0, β_l, and β_k, and the non-parametric functions ϕ_t and $g_{t,t+1}$. The time-series model includes the aggregate shocks ν_t and the parameters ρ that define their distribution function. The decomposition in the two models highlights two features of the example. First, the time-series model affects the cross-sectional counterpart only through the functions ϕ_t and $g_{t,t+1}$. Second, to consistently estimate the production function parameters β_l and β_k, the functions ϕ_t and $g_{t,t+1}$ must be known to control for the correlation between labor and capital on one side and the productivity shocks on the other. But it is irrelevant how the aggregate shocks and the corresponding parameters enter those functions. These two features imply that, if the econometrician is only interested in estimating the production function parameters β_l and β_k, he can achieve this by simply estimating the cross-sectional model. This is possible as long as the functions ϕ_t and $g_{t,t+1}$ are allowed to vary in a non-parametric way over time to deal with the existence of the aggregate shocks. The clever use of time-dummies, therefore, solves all the issues raised by the presence of aggregate shocks. However, if the econometrician is interested in estimating the entire model, which includes the parameters that describe the distribution of the aggregate shocks, he has to rely on the general approach based on the combination of cross-sectional and time-series variables.

5 Example 3: A General Equilibrium Model

In this section, we consider as a third example a general equilibrium model of education and labor supply decisions in which aggregate shocks influence individual choices. This example provides additional insights into the effects of aggregate shocks on the estimation of model parameters. Aggregate shocks are particularly important in the estimation of general equilibrium models (Lee and Wolpin (2006), Dix-Carneiro (2014)).[13] Differently from the portfolio and production function examples, in a general equilibrium context the relationship between the cross-sectional and time-series models is generally bi-directional: the cross-sectional parameters cannot be identified from cross-sectional data without knowledge of the time-series parameters and the time-series parameters cannot be identified from time series data without knowing the cross-sectional parameters. Thus, the simple two step procedure employed in the asset pricing example cannot be used here. Instead, simultaneous estimation of time series and cross-sectional parameters is required. Because of this, the general equilibrium example can be employed to illustrate how the inferential and estimation method developed in the paper can be applied to more general cases. We also use the example to show that, in the presence of aggregate shocks, the limiting distribution of the estimator takes generally a more complex mixed normal form instead of the conventional normal form we find in the asset pricing example. Lastly, the example represents the basis of the Monte Carlo exercise we perform later in the paper to determine the ability of our method to account for aggregate shocks.

In principle, we could have used as a general example a model proposed in the general equilibrium literature such as the model developed in Lee and Wolpin (2006). We decided against this alternative because in those models the effect of the aggregate shocks on the estimation of the model parameters and the relationship between the cross-sectional and time-series models are complicated and therefore difficult to describe. Instead, we decided to develop a model that is sufficiently general to generate an interesting relationship between the shocks and the estimation of the parameters in the cross-sectional and time-series models, but at the same time sufficiently stylized for these relationships to be easy to describe and understand.

[13] Other papers have estimated general equilibrium models without including aggregate shocks. Some examples are Davidson and Woodbury (1993), Ferreyra (2007), Lise, Seitz, and Smith (2004), Metha (2017), and Shephard (2017).

In the model we develop, aggregate shocks affect the education decisions of young individuals and their subsequent labor supply decisions when of working-age. Specifically, we consider an economy where in each period t a young and a working-age generation overlap. Each generation is composed of a continuum of individuals with measure N_t.[14] Each individual is endowed with preferences over a non-durable consumption good and leisure. The preferences of individual i are represented by a Cobb-Douglas utility function $U^i(c,l) = \left(c^{\sigma}l^{1-\sigma}\right)^{1-\gamma_i} \big/ (1-\gamma_i)$, where the risk aversion parameter γ_i is a function of the observable variables $x_{i,t}$, the unobservable variables $\xi_{i,t}$, and a vector of parameters μ, i.e. $\gamma_i = \gamma\left(x_{i,t}, \xi_{i,t} \middle| \mu\right)$. Future utilities are discounted using a discount factor δ.

Both young and working-age individuals are endowed with a number of hours $\mathcal{T}$ that can be allocated to leisure or to a productive activity. Young individuals are also endowed with an exogenous income $y_{i,t}$. In each period, the economy is hit by an aggregate shock ν_t whose conditional probability $P\left(\nu_{t+1} \middle| \nu_t\right)$ is determined by $\log \nu_{t+1} = \varrho \log \nu_t + \eta_t$. We assume that η_t is normally distributed with mean 0 and variance ω^2. The aggregate shock affects the labor market in a way that will be established later on.

In each period t, young individuals choose the type of education to acquire. They can choose education F that trains them for an occupation that is only marginally affected by aggregate shocks. Or they can select education R that prepares them for an occupation that is significantly affected by aggregate shocks. We will refer to education F as the flexible education and to education R as the rigid education. We interpret the flexible education as a an education that provides workers with skills that are valued by firms during periods of economic expansion as well as periods of economic downturn. One example is training that prepares workers for a job in the health or education sectors. The rigid education, instead, endows individuals with skills that are highly valued during periods of economic growth, but are in limited demand during years of economic decline. The typical example is training for jobs in the financial sector.

The two types of education have identical cost $C_e < y_{i,t}$ and need the same amount of time to acquire $\mathcal{T}_e < \mathcal{T}$. Since young individuals typically have limited financial wealth, we assume that there is no saving decision when young and that any transfer from parents or relatives is included in

[14] In the rest of the section we use interchangeably the word 'measure' and the more intuitive but less precise word 'number' to refer to N_t or similar objects.

non-labor income $y_{i,t}$. We also abstract from student loans and assume that all young individuals can afford to buy one of the two types of education. As a consequence, a young individual will consume the part of income $y_{i,t}$ that is not spent on education.

At each t, working-age individuals draw a productivity shock $\varepsilon_{i,t}^{S}$, for $S = F, R$, which determines how productive their hours of work are in case they choose to supply labor. We assume that the productivity shock is unknown to the individuals when young. They also draw a wage offer. Given the productivity shock and the wage offer, working-age individuals choose how much to work $h_{i,t}$ and how much to consume. If a working-age individual decides to supply $h_{i,t}$ hours of work, the effective amount of labor hours supplied is given by $\exp\left(\varepsilon_{i,t}^{F}\right) h_{i,t}$ for the flexible type of education F and by $\exp\left(\varepsilon_{i,t}^{R}\right) h_{i,t}$ for the rigid type of education R. We assume that $\varepsilon_{i,t}^{S}$ is normally distributed with mean μ_{ε}^{S} and variance σ_S^2, for $S = F, R$, and that $\sigma_F^2 < \sigma_R^2$. To simplify the analysis we normalize $E\left[\exp\left(\varepsilon_{i,t}^{S}\right)\right] = 1$, for $S = F, R$.[15] Throughout the section we will use two definitions of wages. We will denote by w_t^S, for $S = F, R$, the wage per unit of effective labor, and by $w_{i,t}^{S} = w_t^S \exp\left(\varepsilon_{i,t}^{R}\right)$ the actual wage received by the worker for each unit of labor hours $h_{i,t}$.

The economy is populated by two types of firms to whom the working-age individuals supply labor. The first type of firm employs only workers with education F, whereas the second type of firm employs only workers with education R. Both use the same type of capital K, which is assumed to be fixed over periods. The labor demand functions of the two types of firms are assumed to take the following form:

$$\log H_t^{D,F} = \alpha_0 + \alpha_1 \log w_t^F,$$

and

$$\log H_t^{D,R} = \alpha_0 + \alpha_1 \log w_t^R + \log \nu_t,$$

where $H^{D,S}$ is the total demand for *effective labor*, with $S = F, R$, $\alpha_0 > 0$, and $\alpha_1 < 0$. We assume that the two labor demands have identical intercepts and slopes for simplicity.[16] These two labor

[15]The assumption $E\left[\exp\left(\varepsilon_{i,t}^{S}\right)\right] = 1$ implies that $\mu_{\varepsilon}^{S} = -\sigma_S^2/2$, for $S = F, R$.

[16]The labor demand function of the flexible firm can be derived from a Cobb-Douglas production function that is independent of the aggregate shock, i.e. $q_t = \left(H_t^F\right)^{\delta}\left(\bar{K}\right)^{\gamma}$, where $\bar{K}$ is the fixed amount of capital employed by the firm. The labor demand function of the rigid firm can be derived from a Cobb-Douglas production function that depends multiplicately on the aggregate shock, i.e. $q_t = \nu_t^{1-\delta}\left(H_t\right)^{\delta}\left(\bar{K}\right)^{\gamma}$. With these production functions, $\alpha_0 = \left(\log\delta + \gamma\log\bar{K}\right)/(1-\delta)$ and $\alpha_1 = -1\,/(1-\delta)$.

demand functions enable us to capture the idea that workers with more flexible education are affected less by aggregate shocks such as business cycle shocks. The wage for each education group is determined by the equilibrium in the corresponding labor market. It will therefore generally depend on the aggregate shock.

We conclude the description of the model by pointing out that there is only one source of uncertainty in the economy, the aggregate shock, and two sources of heterogeneity across individuals, the risk aversion parameter and the productivity shock.

The problem solved in period t by individual i of the young generation is to choose consumption, leisure, and the type of education that satisfy:

$$\max_{c_{i,t},l_{i,t},c_{i,t+1},l_{i,t+1},S} \frac{\left(c_{i,t}^{\sigma} l_{i,t}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} + \delta \int \frac{\left(c_{i,t+1}^{\sigma} l_{i,t+1}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} dP\left(\nu_{t+1} | \nu_t\right) \tag{18}$$

$$s.t. \quad c_{i,t} = y_{i,t} - C_e \quad \text{and} \quad l_{i,t} = \mathcal{T} - \mathcal{T}_e$$

$$c_{i,t+1} = w_{t+1}^{S}\left(\nu_{t+1}\right) \exp\left(\varepsilon_{i,t+1}^{S}\right)\left(\mathcal{T} - l_{i,t+1}\right) \quad \text{for every } \nu_{t+1}.$$

Here, $w_{t+1}^{S}\left(\nu_{t+1}\right)$ denotes the wage per unit of effective labor in the second period, which varies with the education choice $S = F, R$. It is determined in equilibrium and, hence, it depends on the realization of the aggregate shock ν_{t+1}.

The problem solved by a working-age individual takes a simpler form. Conditional on the realization of the aggregate shock ν_t, the individual idiosyncratic shock, and the type of the education S chosen when young, individual i of the working-age generation chooses consumption and leisure that solve the following problem:

$$\max_{c_{i,t},l_{i,t}} \frac{\left(c_{i,t}^{\sigma} l_{i,t}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} \tag{19}$$

$$s.t. \; c_{i,t} = w_{t}^{S}\left(\nu_{t}\right) \exp\left(\varepsilon_{i,t}^{S}\right)\left(\mathcal{T} - l_{i,t}\right).$$

We now solve the model starting from the problem of a working-age individual. Using the first order conditions of problem (19) the optimal choice of consumption, leisure, and hence labor supply for a working-age individual takes the following form:

$$c_{i,t}^* = \sigma w_t^S (\nu_t) \exp\left(\varepsilon_{i,t}^S\right) \mathcal{T}, \tag{20}$$

$$l_{i,t}^* = (1-\sigma)\mathcal{T}, \tag{21}$$

$$h_{i,t}^* = \mathcal{T} - l_{i,t}(\nu_{i,t}) = \sigma \mathcal{T}.$$

The supply of effective labor is therefore equal to $\sigma \exp\left(\varepsilon_{i,t}^S\right)\mathcal{T}$. Given the optimal choice of consumption and leisure, conditional on the aggregate shock, the value function of a working-age individual with education S can be written as follows:

$$V_{i,t}(S, \nu_t, \varepsilon_{i,t}) = \frac{\left[\left(\sigma w_t^S(\nu_t)\exp\left(\varepsilon_{i,t}^S\right)\mathcal{T}\right)^{\sigma}\left((1-\sigma)\mathcal{T}\right)^{1-\sigma}\right]^{1-\gamma_i}}{1-\gamma_i}, \quad \text{S = F, R.}$$

Given the value functions of a working-age individual, we can now characterize the education choice of a young individual. This individual will choose education F if the expectation taken over the next period aggregate shocks of the corresponding value function is greater than the analogous expectation for education R:

$$E\left[V_{i,t}(F, \nu_{t+1}, \varepsilon_{i,t+1}) \middle| \nu_t\right] \geq E\left[V_{i,t}(R, \nu_{t+1}, \varepsilon_{i,t+1}) \middle| \nu_t\right]. \tag{22}$$

To simplify the discussion, we assume that $\varepsilon_{i,t+1}$ is independent of γ_i, thereby eliminating sample selection issues in the wage equations.

Before we can determine which variables and parameters affect the education choice, we have to derive the equilibrium in the labor market. It can be shown that the labor market equilibrium is characterized by the following two wage equations:[17]

$$\log w_{i,t}^F = \frac{\log n_t^F + \log\sigma + \log\mathcal{T} - \alpha_0}{\alpha_1} + \varepsilon_{i,t}^F, \tag{23}$$

$$\log w_{i,t}^R = \frac{\log n_t^R + \log\sigma + \log\mathcal{T} - \alpha_0 - \log\nu_t}{\alpha_1} + \varepsilon_{i,t}^R, \tag{24}$$

where $w_{i,t}^F$ and $w_{i,t}^R$ are the wages individual i would receive if i chooses to work in sector F or

[17]See the on-line Appendix B.2.

R and n_t^F and n_t^R are the measures of individuals that choose education F and R. We can now replace the equilibrium wages inside inequality (22) and analyze the education decision of a young individual. It can be shown that a young individual chooses the flexible type of education at time t if the following inequality is satisfied:[18]

$$\gamma_i \geq 1 - \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}}. \tag{25}$$

This inequality provides some insight into the educational choices of young individuals. Since $\alpha_1 < 0$, they are more likely to choose the flexible education which insures them against aggregate shocks if the variance of the aggregate shock is larger, if they are more risk averse, if the aggregate shock at the time of the decision is lower as long as $\varrho > 0$, and if the elasticity of the wage for the rigid education with respect to the aggregate shock is larger (the absolute value of α_1 is lower).

Similarly to the first two examples, we can classify some of the variables and some of the parameters as belonging to the cross-sectional model and the remaining to the time-series model. The cross-sectional variables include consumption $c_{i,t}$, leisure $l_{i,t}$, individual wages $w_{i,t}^F$ and $w_{i,t}^R$, the variable determining the educational choice $D_{i,t}$, the amount of time $\mathcal{T}$ an individual can divide between leisure and productive activities, and the variables that enter the risk aversion parameter $x_{i,t}$. The time-series variables are composed of the aggregate shock ν_t, the numbers of young individuals choosing the two types of education n^F and n^R, and the aggregate equilibrium wages in the two sectors $w_t^F = E\left[w_{it}^F\right]$ and $w_t^R = E\left[w_{it}^F\right]$.[19] We want to stress the difference between individual wages and aggregate wages. Individual wages are typically observed in panel data or repeated cross-sections whose time dimension is generally short, whereas aggregate wages are available in longer time-series of aggregate data. The cross-sectional parameters consist of the relative taste for consumption σ, the variances σ_F^2 and σ_R^2 of the individual productivity shocks, the parameters defining the risk aversion μ, and the parameters of the wage equations α_0 and α_1, whereas the time-series parameters include the two parameters governing the evolution of the aggregate shock ϱ and ω^2, and the discount factor δ. The discount factor is notoriously difficult

[18]Details are given in the on-line Appendix.

[19]The expectation operator E corresponds to the expectation taken over the distribution of cross-sectional variables.

to estimate. For this reason, in the rest of the section we will assume it is known.

We now employ the method proposed in this paper, which exploits a combination of a long time-series of aggregate data and cross-sectional data, in the estimation of the model parameters. We assume that the econometrician has access to two repeated cross-sections of data for periods $t = 1$ and $t = 2$, which include i.i.d. observations on educational choices $F_{i,t}$, wages $w_{i,t}^S$ with $S = F, R$, consumption $c_{i,t}^*$, and leisure $l_{i,t}^*$. The econometrician also has access to a time-series of aggregate data that spans $s = \tau_0 + 1, \ldots, \tau_0 + \tau$. It consists of the measures of people choosing the flexible and rigid educations n_t^F and n_t^R, and their corresponding aggregate wages w_s^F and w_s^R. For simplicity, we assume that the two cross-sections consist of the same number of individuals n, and that the first $\bar{n}_1$ and $\bar{n}_2$ individuals in the two cross sections chose $S = F$.

The parameters α_1, σ, σ_F^2, and σ_R^2 can be estimated using only the two cross-sections. Specifically, α_1 can be consistently estimated using the wage equation for flexible education (23) in periods 1 and 2 as the $\widehat{\alpha}_1$ that solves

$$\frac{1}{\bar{n}_1} \sum_{i=1}^{\bar{n}_1} \log w_{i,1}^F - \frac{1}{\bar{n}_2} \sum_{i=1}^{\bar{n}_2} \log w_{i,2}^F = \frac{1}{\widehat{\alpha}_1} \left(\log n_1^F - \log n_2^F \right). \tag{26}$$

This can be done because the productivity shock ε_t and the risk aversion parameter γ_i are assumed to be independent of each other, which implies that there is no sample selectivity problem. The parameter σ can be consistently estimated employing the consumption and leisure choices of working-age individuals (20) and (21) for period 1 as the $\widehat{\sigma}$ that solves

$$\frac{1}{\bar{n}_1} \sum_{i=1}^{\bar{n}_1} \frac{c_{i,1}^*}{l_{i,1}^*} = w_1^F \frac{\widehat{\sigma}}{1 - \widehat{\sigma}}. \tag{27}$$

The variances of the productivity shocks for the two sectors σ_F^2 and σ_R^2 can be estimated using the wage equations for sectors F and R (23) and (24) as the sample variances of $\log w_{i,t}^F$ and $\log w_{i,t}^R$.

The aggregate shocks and the parameters governing their evolution ϱ and ω^2 can then be estimated using the time-series of aggregate data. Specifically, with α_1 consistently estimated, the aggregate shock in period s can be consistently estimated for $s = \tau_0 + 1, \ldots, \tau_0 + \tau$ using the following equation:

$$\widehat{\log \nu_s} = \widehat{\alpha}_1 \left(\log w_s^F - \log w_s^R \right) - \left(\log n_s^F - \log n_s^R \right), \tag{28}$$

which was derived by computing the difference between the equations defining the equilibrium wages in sectors R and S and solving for $\log \nu_s$.[20] Observe that ν_s can only be estimated because α_1 was previously estimated using the cross-sections. The parameters ϱ and ω^2 can then be consistently estimated by the time-series regression of the equation that characterizes the evolution of the estimated aggregate shocks:

$$\widehat{\log \nu_{s+1}} = \varrho \widehat{\log \nu_s} + \eta_t. \tag{29}$$

The only parameters left to estimate are the parameters μ defining the individual risk aversion γ_i. They are the most interesting parameters of the model because they incorporate the bi-directional relationship between the cross-sectional and time-series models, as the following discussion reveals. Specifically, if the distribution of γ_i is parametrically specified, the parameters μ can be consistently estimated by MLE using cross-sectional variation on the educational choices and the inequality that characterizes those choices (25). In the Monte Carlo exercise in Section 7, we assume that $\log \gamma_i \sim N(\mu, 1)$. Under this assumption, the distribution of risk aversion in the population is characterized by only one parameter, its mean μ. It can be shown that in this case the probability that an individual chooses education F takes the following form:[21]

$$1 - \Phi\left(\log\left(1 - \Theta_t\right) - \mu\right),$$

where Φ denotes the CDF of $N(0,1)$, and

$$\Theta_t \equiv \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}}. \tag{30}$$

We can therefore estimate the mean of the distribution of risk aversion μ using a Probit maximum likelihood estimator, provided that ν_t, ϱ, ω^2, σ_F^2, σ_R^2, σ, and α_1 and are known.[22] The cross-sectional parameter μ can therefore be estimated only if the time-series parameters ν_t, ϱ, and ω^2

[20] The equations defining the equilibrium wages are reported in the Appendix as equations (52) and (53).

[21] For details see the on-line Appendix.

[22] It is straightforward to relax the distributional assumption on γ_i and consider the more general case where the risk aversion parameter γ_i is a function of the observable variables $x_{i,t}$, the unobservable variables $\xi_{i,t}$, and a vector of parameters μ, i.e. $\gamma_i = \gamma\left(x_{i,t}, \xi_{i,t} \middle| \mu\right)$.

have been previously estimated. But their estimation requires the prior estimation of the cross-sectional parameter α_1. Hence, the bi-directional relationship between the cross-sectional and time-series models.

To evaluate the effect of misspecification by ignoring aggregate shocks when estimating the parameters of the general equilibrium model, we now consider the case of an econometrician who is unaware of the presence of aggregate shocks and, hence, only uses cross-sectional variation for the identification and estimation of the parameters of interest. The misspecification changes the inequality that characterizes the education choice (25), which in this case takes the following form:[23]

$$\gamma_i \geq 1 - \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \log \nu_{t+1}}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2\right)}{2\alpha_1}}. \tag{31}$$

As a consequence, under the misspecification and the assumption that $\log \gamma_i \sim N(\mu, 1)$, the probability that an individual chooses education F becomes

$$1 - \Phi\left(\log\left(1 - \Theta_t^*\right) - \mu\right),$$

where

$$\Theta_t^* \equiv \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \log \nu_{t+1}}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2\right)}{2\alpha_1}}.$$

Since this form of misspecification only changes the probability of choosing education F, only estimation of the parameter μ is affected. To understand its effect, we derive the estimation bias in closed form. In the misspecified model, the probability that someone selects education F can be written as follows:

$$1 - \Phi\left(\log\left(1 - \Theta_t^*\right) - \mu\right) = 1 - \Phi\left(\log\left(1 - \Theta_t\right) - \left(\mu - \log\left(1 - \Theta_t^*\right) + \log\left(1 - \Theta_t\right)\right)\right).$$

Let $\hat{\mu}$ be the maximum likelihood estimator of the correctly specified model. Then, the previous equation implies that the maximum likelihood estimator $\widehat{\mu}_{\text{mis}}$ of the misspecified model satisfies

[23]For details see the on-line Appendix E.

the following equation:

$$\widehat{\mu}_{\text{mis}} = \widehat{\mu} + \log\left(1 - \hat{\Theta}_t^*\right) - \log\left(1 - \hat{\Theta}_t\right),$$

where $\hat{\Theta}_t^*$ and $\hat{\Theta}_t$ denote the estimators of Θ_t^* and Θ_t. The asymptotic misspecification bias has therefore the following analytic form:

$$\log\left(1 - \Theta_t^*\right) - \log\left(1 - \Theta_t\right) = \tag{32}$$

$$\log\left(1 - \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \log\nu_{t+1}}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2\right)}{2\alpha_1}}\right) - \log\left(1 - \frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho\log\nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}}\right).$$

It shows that the magnitude of the asymptotic bias depends on the size of the the variance of the aggregate shocks ω^2 and on the difference between the expected aggregate shock in period $t+1$, $\varrho\log\nu_t$, and its realization, $\log\nu_{t+1}$. Later in the paper, we will use particular values for the model parameters to provide evidence on the magnitude of the bias.

Intuitively, ignoring the uncertainty generated by the aggregate shocks should have the same effect as in the portfolio example of biasing upward the estimated risk aversion parameter. Not accounting for the aggregate shocks is equivalent to assuming that the agents face less uncertainty than they actually experience when making the education decisions. Since the individuals' decisions are based on the actual uncertainty, the only way the model can explain those choices is by making people more risk averse. In the general equilibrium model, this insight is not as straightforward to see as in the portfolio example, since the bias depends also on the difference between the current and next period aggregate shocks. For this reason we perform a Monte Carlo exercise whose results are reported in Section 7. They confirm the intuition regarding the sign of the bias and suggests that its size can be extremely large. These insights are not specific to the uncertainty generated by the aggregate shocks. They apply equally to individual-specific shocks. If the econometrician disregards the variation generated by those shocks, risk aversion will generally be estimated to be larger than it actually is.

There is an alternative approach that uses only micro-level data, instead of a combination of micro-level and aggregate data, to estimate model parameters when aggregate shocks affect behavior. The econometrician can use a single panel of micro-level data in which the time-series dimension of the panel is sufficiently long, instead of a small number of repeated cross-sections combined with the time-series of aggregate data. The general equilibrium model of this section is too complicated to illustrate the limitations of the alternative panel-data approach. Let n and $\mathbb{T}$ denote the cross section dimension and time series dimension of the panel data. Using a stylized linear panel model, however, one can show that, when the alternative approach is used, the effective sample size of the data is not $n \times \mathbb{T}$ but $\mathbb{T}$, with the cross-section generally playing a minor role.[24] The reason is that the asymptotic theory for the alternative "long panel" approach requires, analogously to time-series analysis, the time dimension $\mathbb{T}$ to go to infinity because parameters related to the aggregate shock process are exclusively identified from time series variation. A large cross-section n does not compensate for the lack of a long time-series in the panel. Since in practice almost all panel data sets have limited time-series dimensions, using the alternative panel approach would therefore lead to imprecise estimates relative to our proposed method.

It is also important to point out that the practice of computing standard errors under the assumption that the time-series parameters are known does not solve the large-$\mathbb{T}$ problem illustrated by our panel example. Under that assumption, the standard errors for the cross-sectional parameters are incorrect and tend to be too small because they do not account for the noise introduced by the estimation of the time series parameters. Lee and Wolpin (2006) use such a procedure (see also their footnote 37). Their standard errors therefore underestimate the true standard errors.[25]

The econometric method proposed in this paper for the estimation of models with aggregate shocks requires the combined use of cross-sectional data with long time-series of aggregate data. There are no formulas available for the computation of standard errors and confidence intervals that account for jointly estimated time series and cross-sectional coefficients based on those combined data sources. In the next section, we provide such formulas. They are based on a new and complex asymptotic theory that we develop in the companion paper Hahn, Kuersteiner, and Mazzocco (2016). Surprisingly, in spite of the complexity of the theory, the formulas are straightforward and easy to use.

[24]A detailed exposition of the model and derivation are in the on-line Appendix C.

[25]Donghoon Lee kindly confirmed this in private communication.

6 Standard Errors

The asymptotic theory underlying estimators obtained from the combination of the two data sources considered in this paper is complex. It is based on a new central limit theorem that requires a novel martingale representation. Given its complexity, the theory is presented in a separate paper (Hahn, Kuersteiner and Mazzocco (2016)). However, the mechanical implementation of test statistics and confidence intervals is surprisingly straightforward. In this Section, we first provide a step-by-step description of how these statistics can be calculated. We then explain how they can be employed in concrete cases using as examples the portfolio choice and the general equilibrium models analyzed in the previous sections.

The computation starts with the explicit characterization of the "moments" that identify the cross-sectional parameters β and the time-series parameters ρ. In the most general case, the aggregate shocks are unknown and must be estimated jointly with the other model parameters using cross-sectional data, as illustrated in the general equilibrium example. The shocks can therefore be treated as cross-sectional parameters. This is accounted for by introducing a new vector of parameters θ which is composed of the original cross-sectional parameters and the aggregate shocks, i.e. $\theta = (\beta, \nu_1, ..., \nu_T)$.[26] We then denote with $f_{\theta,i}(\theta, \rho)$ the i-th moment used in the identification of the parameters in θ and with $g_{\rho,t}(\beta, \rho)$ the t-th moment used in the identification of the time-series parameters. For simplicity, we assume that $g_{\rho,t}$ is the correctly specified score of the conditional density of z_t. Since the score of a correctly specified likelihood is a martingale difference sequence it is not serially correlated. This simplifies the estimation of Ω_g.[27]

Our proposed estimator based on a combination of cross-sectional data and a long time-series

[26] Implicit in this representation is the idea that we are given a short panel for estimation of $\theta = (\beta, \nu_1, \ldots, \nu_T)$, where T denotes the time series dimension of the panel data. In order to emphasize that T is small, we use the term 'cross-section' for the short panel data set, and adopt asymptotics where T is fixed.

[27] In a technical note we discuss extensions to the misspecified and non-martingale case. There we provide results for versions of our standard errors that are robust to serial correlation. More specifically, we prove a joint time series and cross-sectional CLT for stationary and iid processes allowing for the case where $E[g_{\rho,t}] = 0$ but not necessarily $E[g_{\rho,t}g'_{\rho,s}] = 0$ for $t \neq s$. It follows that $\Omega_g = E[g_{\rho,t}g'_{\rho,t}] + \sum_{s=1}^{\infty}\left(E[g_{\rho,t}g'_{\rho,s}] + E[g_{\rho,s}g'_{\rho,t}]\right)$ which can be estimated consistently with a HAC standard error estimator. The remaining implications of our theory as presented in this paper are unaffected by this extension.

of aggregate data can then be written as the solution $\left(\hat{\theta}, \hat{\rho}\right)$ to the following system of equations:

$$\sum_{i=1}^{n} f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right) = 0, \tag{33}$$

$$\sum_{s=\tau_0+1}^{\tau_0+\tau} g_{\rho,s}\left(\hat{\beta}, \hat{\rho}\right) = 0. \tag{34}$$

As discussed in Section 2, $\tau_0 + 1$ here denotes the beginning of the time series data, which is allowed to differ from the beginning of the panel data.

Using those equations, the standard errors for $\hat{\theta}$ and $\hat{\rho}$ can be calculated using the following five steps.

1. Let $\phi = (\theta', \rho')'$ be the vector of parameters.

2. Let

$$\mathbf{A} = \begin{bmatrix} \hat{A}_{f,\theta} & \hat{A}_{f,\rho} \\ \hat{A}_{g,\theta} & \hat{A}_{g,\rho} \end{bmatrix},$$

   be the matrix of first order derivatives of the moments with respect to the parameters, with

$$\hat{A}_{f,\theta} = n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)}{\partial \theta'}, \qquad \hat{A}_{f,\rho} = n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)}{\partial \rho'},$$

$$\hat{A}_{g,\theta} = \tau^{-1} \sum_{s=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,s}\left(\hat{\beta}, \hat{\rho}\right)}{\partial \theta'}, \qquad \hat{A}_{g,\rho} = \tau^{-1} \sum_{s=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,s}\left(\hat{\beta}, \hat{\rho}\right)}{\partial \rho'}.$$

3. Let

$$\hat{\Omega}_f = \frac{1}{n} \sum_{i=1}^{n} f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right) f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)'$$

   and

$$\hat{\Omega}_g = \frac{1}{\tau} \sum_{s=\tau_0+1}^{\tau_0+\tau} g_{\rho,s}\left(\hat{\theta}, \hat{\rho}\right) g_{\rho,s}\left(\hat{\theta}, \hat{\rho}\right)'.$$

4. Let

$$W = \begin{bmatrix} \frac{1}{n}\hat{\Omega}_f & 0 \\ 0 & \frac{1}{\tau}\hat{\Omega}_g \end{bmatrix}$$

5. Calculate

$$\mathbf{V} = \mathbf{A}^{-1} W \left(\mathbf{A}'\right)^{-1}$$

and use the square roots of the diagonal elements as the standard errors of the estimator. For instance, if one is interested in the 95% confidence interval of the first component of ϕ, it can be written as $\hat{\phi}_1 \pm 1.96\sqrt{\mathbf{V}_{1,1}}$.

The five-step algorithm described above may be understood intuitively by considering a Taylor series expansion of equations (33) and (34). For simplicity, we work with a special case where the time series model does not depend on the cross-section parameter β and the derivative with respect to ρ of the log of the probability density function of the time series variable z_s conditional on z_{s-1} is given by $g_\rho\left(z_s|\, z_{s-1}, \rho\right)$. Denote by $\hat{\rho}$ the MLE. Standard results imply that $\sqrt{\tau}\left(\hat{\rho} - \rho\right)$ is asymptotically $N\left(0, A_{g,\rho}^{-1}\Omega_g\left(A_{g,\rho}'\right)^{-1}\right)$, where $A_{g,\rho} = E\left[\partial g_{\rho,t}/\,\partial\rho'\right]$ and $\Omega_g = E\left[g_{\rho,s}g_{\rho,s}'\right]$ with $g_{\rho,s} = g_\rho\left(z_s|\, z_{s-1}, \rho\right)$. The variance matrix of $\hat{\rho}$ therefore corresponds to the lower-right block of the matrix $\mathbf{V}$ derived using the five-step algorithm.

With $\hat{\rho}$ estimated from the time series data, we can estimate θ by solving (33). Using a Taylor series expansion, the asymptotic distribution of $\hat{\theta}$ is characterized as follows:

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \approx -\left(n^{-1}\sum_{i=1}^{n}\frac{\partial f_{\theta,i}\left(\theta,\rho\right)}{\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{\theta,i}\left(\theta,\hat{\rho}\right)\right).$$

Because

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{\theta,i}\left(\theta,\hat{\rho}\right) \approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{\theta,i}\left(\theta,\rho\right) + \frac{\sqrt{n}}{\sqrt{\tau}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f_{\theta,i}\left(\theta,\rho\right)}{\partial\rho'}\right)\sqrt{\tau}\left(\hat{\rho}-\rho\right),$$

we expect

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \approx -A_{f,\theta}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{\theta,i}\right) - A_{f,\theta}^{-1}A_{f,\rho}\frac{\sqrt{n}}{\sqrt{\tau}}\sqrt{\tau}\left(\hat{\rho}-\rho\right), \qquad (35)$$

where $A_{f,\theta} = E\left[\partial f_{\theta,i}/\,\partial\theta'\right]$ and $A_{f,\rho} = E\left[\partial f_{\theta,i}/\,\partial\rho'\right]$. Assume that $n^{-1/2}\sum_{i=1}^{n} f_{\theta,i}$ is asymptotically $N\left(0,\Omega_f\right)$, and that n and τ grow to infinity at the same rate. If the two terms on the right hand side of (35) are asymptotically independent, we expect $\sqrt{n}\left(\hat{\theta}-\theta\right)$ to be approximately $N\left(0, A_{f,\theta}^{-1}\Omega_f\left(A_{f,\theta}'\right)^{-1} + \frac{n}{\tau}A_{f,\theta}^{-1}A_{f,\rho}A_{g,\rho}^{-1}\Omega_g\left(A_{g,\rho}'\right)^{-1}A_{f,\rho}'\left(A_{f,\theta}'\right)^{-1}\right)$ or, equivalently, we expect $\hat{\theta}-\theta$ to be approximately normally distributed with variance equal to $\frac{1}{n}A_{f,\theta}^{-1}\Omega_f\left(A_{f,\theta}'\right)^{-1} +$

$\frac{1}{\tau}A_{f,\theta}^{-1}A_{f,\rho}A_{g,\rho}^{-1}\Omega_g\left(A'_{g,\rho}\right)^{-1}A'_{f,\rho}\left(A'_{f,\theta}\right)^{-1}$. This is exactly the variance of $\hat{\theta}$ implied by the upper-left block of the matrix $\mathbf{V}$ derived using the five-step algorithm. In general, the two terms on the right hand side of (35) are not asymptotically independent. But our technical discussion in the on-line appendix justifies this assumption.

The theoretical results in our companion paper as well as more detailed calculations in the appendix reveal a few important points. The matrix $\mathbf{V}$ in general is a function of aggregate shocks realized during the observation periods of the cross-sectional sample. Consequently, the standard errors computed from a combination of cross-sectional and time series data generally depend on the actual realizations of an aggregate shocks at the time the cross-sections were observed. With variation across the business cycle, these shocks and therefore the estimated standard errors may change. Explicit formulas of how the standard errors depend on aggregate shocks in our general equilibrium model are given in the appendix. The consequence of this finding is that comparing standard errors across studies with cross-sections observed at different points in time is problematic. This result applies to both cross-sectional and time-series parameters. A similar word of caution applies to sample descriptive statistics such as sample averages obtained from short panels, since these averages in general are functions of realized values of aggregate shocks even when the cross-sectional sample size is large. As a result, descriptive statistics are expected to change in response to changes of the aggregate shock. Comparison of these descriptive measures across different time periods or data sets thus needs to be done with caution. Pivotal statistics such as t-ratios or confidence intervals have, however, standard distributional properties and can be compared across different samples. The deep structural parameters estimated in this paper are also typically thought to be fixed. As long as these parameters are estimated consistently, their point estimators are not affected by variation from aggregate shocks in large enough samples.

In Appendix D of the on-line appendix, we show for the interested reader how the standard error formulas can be derived for the portfolio example of Section 3 and the general equilibrium model of Section 5. The application of the formulas to the two examples highlights two features that determine the properties of the asymptotic distribution of the proposed estimator. In the simple portfolio example, there is a unidirectional relationship between the cross-sectional and time-series models. As a consequence, the cross-sectional parameters cannot be estimated without knowledge of the time-series parameters. In addition, agents form expectations for the main variable, end-of-period wealth, that do not depend on the current realization of the aggregate shock. These two features imply that the asymptotic distribution has a simple form that is independent of the aggregate shocks. If one of these two conditions is not satisfied, the limiting distribution has the more general and complicated form that depends on aggregate shocks. The more complex general equilibrium example illustrates this point. In that case, the relationship between the two sub-models is bidirectional, implying that there is no recursive structure that can be used to first estimate the cross-sectional parameters without knowledge of their time-series counterparts. As a consequence, the asymptotic distribution depends on the aggregate variables needed for the estimation of the cross-sectional parameters. Moreover, agents use the current realization of the aggregate shock to form expectations about future events. Since these expectations are used in their decision making process, the aggregate shocks affect the limiting distribution of the estimator by entering the variance-covariance matrix.

7 Monte Carlo Results

In this section, we present the Monte Carlo results obtained by simulating the general equilibrium model. We use the Monte Carlo results first to illustrate how the estimation and inference approach developed in this paper can be applied in practice. We then document the ability of our standard error formulas to produce the correct coverage probabilities for the parameters of interest, and the inability of the standard error formulas that do not account for the aggregate uncertainty to generate appropriate coverage probabilities. As a by-product, we also report the magnitude of the estimation bias that can be generated if the econometrician ignores aggregate shocks.

To perform the Monte Carlo simulations and determine the size of the bias, we have to set the 7 parameters of the general equilibrium model at particular values. The most consequential parameter value is the one assigned to the variance of the aggregate shocks ω^2 since, as shown in Section 5, it determines the magnitude of the bias if the econometrician ignores the aggregate shocks. We chose the size of ω^2 using the estimated variance of the aggregate shocks used by Kydland and Prescott (1982). They use an estimated variance for the quarterly U.S. cyclical output that is equal to 0.000165. Differently from Kydland and Prescott (1982), in our model capital is assumed to be fixed. As a consequence, the variation in aggregate shocks affects exclusively labor demand. To account for this feature of our model, we divided the variance estimated in Kydland and Prescott (1982) by the square of the labor share in the economy.[28] Since in the U.S. the labor share is approximately 2/3, we divide 0.000165 by 4/9 to obtain 0.00037. In addition, our model is characterized by two sectors: the rigid and the flexible sectors. We interpret the flexible sector as the one composed of workers with jobs in industries that are affected less by aggregate shocks (e.g. health and education), and the rigid sector as the one that employs workers with jobs in more cyclical sectors (e.g. finance and business). If one includes in the rigid sector financial activities, professional and business services, and construction, the Bureau of Labor Statistics (BLS) estimates that slightly less than one quarter of workers where employed in the rigid sector (22.6% in 2006 and 22.5% in 2016).[29] Since in our model only the rigid sector is affected by aggregate shocks, to make the estimated variance consistent with our model, we multiply it by the square of 4 (the inverse of the size of the rigid sector). With this additional adjustment, we have a quarterly variance for the aggregate shock of 0.006. Our model has only two periods, one in which people engage in education and one in which they work. We assume that each period is composed of 20 years and we multiply the quarterly variance of 0.006 by 4 quarters and 20 years, obtaining the aggregate variance we use in the simulations, 0.48.

The values assigned to the variances of the productivity shocks σ_F^2 and σ_R^2 are also important for the outcome of the Monte Carlo exercise, since they determine the size of the individual-level uncertainty relative to the size of the aggregate uncertainty. We chose those variances using the estimated variance of the productivity shocks reported in Macurdy (1982). Macurdy (1982) estimates a variance for the residuals of yearly wages in the U.S. that is between 0.054 and 0.062. To derive our measures of the micro variances, we multiply the upper bound of the yearly variance estimated by Macurdy by 20 years (one of our periods), obtaining 1.2.[30] Lastly, in our model the micro shocks in sector F have a smaller variance than the shocks in sector R. To account for this, we set $\sigma_F^2 = 1$ and $\sigma_R^2 = 1.4$. The mean variance of the micro shocks is therefore 1.2, which corresponds to the estimate obtained using the results in Macurdy (1982).

The remaining parameters are set equal to the following values. The mean of the log of the

[28]The derivation of the short-run labor demand function for a Cobb-Douglas production function shows that this is the correct adjustment.

[29]The BLS data used in the computation are available at $https://www.bls.gov/emp/ep\_table\_201.htm$.

[30]If we use the lower bound, the bias increases.

risk aversion parameter μ is set equal to 0.2, which corresponds to a mean risk aversion parameter of approximately 2. The parameter measuring the persistence of the aggregate shock ρ is initially set equal to 0.75. We then evaluate how the results change when it is first increased to 0.9 and then reduced to 0.5. The constant α_0 and slope α_1 of the labor demand functions are chosen to be equal to 7 and -1, respectively. The parameter characterizing the preferences for consumption σ is set equal to 0.6.

In the Monte Carlo exercise we consider 12 different specifications depending on the size of the cross-section sample and length of the time-series sample. Specifically, we simulate the model and estimate the parameters using the following sample sizes for the cross-section: 2,500, 5,000, 10,000, and 50,000 individuals; and the following lengths for the time-series: 25, 50, and 100 periods. Lee and Wolpin (2006) construct their cross-sections using the National Longitudinal Survey of Youth 1979 (NLSY79), with sample size 12,686 per wave, and the Current Population Survey (CPS) March Supplement, with sample size 50,000 until 2000 and 60,000 afterward. Their time-series is constructed using Bureau of Economic Analysis (BEA) data for the period 1968-2000, which implies that its length corresponds to 33. The specification that approximates Lee and Wolpin's setting is therefore the one with 10,000 individuals in the cross-section and 25 periods in the time-series. In all cases we generate 5,000 simulated data sets for the general equilibrium model. The estimates and the coverage probabilities obtained using the method proposed in this paper and the coverage probabilities computed without considering the aggregate uncertainties are presented in Table 1. The mean and median bias generated by ignoring the aggregate shocks are reported in Table 3. We only report results for the parameters μ, ρ, and ω^2. All the other parameters are estimated using the same estimators in the correct and misspecified model. The estimates are therefore identical in the two models. Moreover, they are estimated precisely and without significant bias in all Monte Carlo specifications.

We start by discussing the performance of the proposed approach. In the second column of Table 1, we report the selected parameter estimates and, in the third and fourth columns, the coverage probability for those parameters of a confidence interval with 90% nominal coverage probability.[31] Table 1 documents that the accuracy of the estimates increases with the length of the time-series. For all cross-sections, when the length of the time-series increases from 25 to 100,

[31] To perform the Monte Carlo exercise we have to deal with a technical issue. The estimation of the risk aversion

the estimated persistence parameter ρ goes from 0.698, 0.052 lower than the true parameter, to about 0.735, just 0.015 lower than the true parameter. The size of the cross-section has therefore no effect on the estimated value of ρ. A similar pattern characterizes the estimates of the variance of the aggregate shocks, except that in this case the size of the cross-section has a small effect on the estimation results. For a cross-section of 10,000 individuals, an increase from 25 to 100 periods produces a decline in the estimated ω^2 from 0.503, 0.023 higher than the true parameter, to 0.487, just 0.007 above the true value. Similar trends characterize the estimates of ω^2 for the other cross-sections, except that the accuracy of the estimates improves slightly for cross-sections larger than 2,500.

In the estimation of the risk aversion parameter μ, we replace the other parameters that enter the educational decision (25) with their estimated values. The small biases in the estimation of ρ and ω^2 will therefore affect the estimation of μ, and generate patterns that are similar to the ones observed for ρ and ω^2 when we increase the length of the time-series and the size of the cross-section. For instance, with a cross-section of 10,000 individuals, when we increase the time-series from 25 to 100 periods the estimated μ increases from 0.173, 0.027 below the true parameter, to 0.184, just 0.016 below the true value. To confirm that the small bias in the estimation of μ is generated by the small biases that characterize the other parameters, we also estimated μ using the educational decision and the true value of the other parameters. We will refer to this estimator as the infeasible estimator. The estimated values obtained using this estimator, which by construction varies only with the length of the time series, are reported in Table 1. They are always identical to the true parameter, which confirms that the small bias in the estimation of μ is generated by the small bias introduced by the other parameters. These results indicate that it

---

parameter μ in the general equilibrium model requires the computation of $\log(1-\Theta_t)$ where

$$\Theta_t \equiv \frac{\log\left(\frac{n^F_{t+1}}{n^R_{t+1}}\right) + \frac{\sigma_R^2-\sigma_F^2}{2} + \varrho \log \nu_t}{\frac{\sigma\left(\sigma_R^2-\sigma_F^2+\omega^2\right)}{2\alpha_1}}.$$

In the model, Θ_t is always smaller than 1 and, hence, $\log(1-\Theta_t)$ is always well defined. We provide a proof of this statement in Appendix G. In the estimation of μ, however, the true parameters included in Θ_t are replaced with their estimated values. In some of the Monte Carlo repetitions, the randomness of the estimated parameters generates values of Θ_t that are greater than 1, which implies that $\log(1-\Theta_t)$ is not well defined. A similar problem arises when we estimate the misspecified model. The results reported in this Section are obtained by dropping all simulations for which $\Theta_t \geq 1$. In Appendix F, we report the results obtained by using all the Monte Carlo runs and by setting $\Theta = 0.99$ in all cases in which $\Theta \geq 1$.

is important to use a long time-series when estimating a model with aggregates shocks to reduce the noise introduced by the estimation of the other parameters. A long time-series of aggregate variables should therefore be preferred to a panel of data, since available panels have a short time dimension.

The second column of Table 1 reports the coverage probabilities using the method we have developed. In all cases, the coverage probabilities are close to the nominal one. For example, with a cross-section of $10,000$ individuals, the coverage probability for a 90% interval is 0.890 for a time-series of 25 periods, 0.913 if the time-series is increased to 50, and 0.928 when 100 periods are used. The third column describes the coverage probabilities obtained using standard errors based on the erroneous assumption that the aggregate parameters are not estimated but known. The coverage probabilities of the erroneous confidence intervals display two noteworthy patterns. The first pattern is that the coverage probabilities move further away from the theoretical one when the size of the cross-section increases. With 100 periods, the erroneous coverage probability is 0.818 when $2,500$ individuals are simulated, but only 0.469 when $50,000$ workers are considered. The explanation for this pattern is straightforward. Under the assumption that the time-series parameters are known, an increase in the size of the cross-section implies that the cross-sectional parameters are estimated with more precision, hence the smaller standard errors. If the assumption is incorrect and the time-series parameters are estimated, the erroneous standard errors incorrectly reflect the true variability of the estimator, often in the direction of smaller confidence intervals. Second, the under-coverage problem is ameliorated as the length of the time series grows. For instance, with a cross-section of $10,000$ people, the coverage probability goes from 0.515 with 25 periods to 0.694 with 100 periods. This pattern can be explained by noting that the erroneous standard errors are computed under the assumption that the parameters of the aggregate shocks are known. When the time series grows, this assumption becomes a better approximation of the economy we simulate and, as a consequence, the under-coverage problem is often reduced.

We now provide a more rigorous explanation for the two patterns discussed above. We start by observing that the erroneous standard errors are obtained using the matrix

$$\frac{1}{n}\hat{A}_{f,\theta}^{-1}\hat{\Omega}_f\left(\hat{A}_{f,\theta}^{-1}\right)', \tag{36}$$

instead of the correct formula $\mathbf{V} = \mathbf{A}^{-1} W (\mathbf{A}')^{-1}$ introduced in the previous section. To simplify the discussion, consider the special case in which the cross-sectional parameter θ is a scalar. In that event, the square of the correct standard error for θ corresponds to the upper left block of $\mathbf{A}^{-1} W (\mathbf{A}')^{-1}$, which can be written as follows:

$$\mathbf{A}^{-1} W (\mathbf{A}')^{-1} = \frac{1}{n} \left(\hat{A}_{f,\theta} - \hat{A}_{f,\rho} \hat{A}_{g,\rho}^{-1} \hat{A}_{g,\theta} \right)^{-1} \hat{\Omega}_f \left(\left(\hat{A}_{f,\theta} - \hat{A}_{f,\rho} \hat{A}_{g,\rho}^{-1} \hat{A}_{g,\theta} \right)^{-1} \right)' + \frac{1}{\tau} \hat{A}_{f,\theta}^{-1} \hat{A}_{f,\rho} \left(\hat{A}_{g,\rho} - \hat{A}_{g,\theta} \hat{A}_{f,\theta}^{-1} \hat{A}_{f,\rho} \right)^{-1} \hat{\Omega}_g \left(\hat{A}_{f,\theta}^{-1} \hat{A}_{f,\rho} \left(\hat{A}_{g,\rho} - \hat{A}_{g,\theta} \hat{A}_{f,\theta}^{-1} \hat{A}_{f,\rho} \right)^{-1} \right)' . \quad (37)$$

An inspection of the previous equation reveals that the incorrect formula (36) ignores the term after the plus sign in (37), which corresponds to the time series error of magnitude $O\left(\frac{1}{\tau}\right)$, whose size may be substantial if n is significantly bigger than τ.

To illustrate the effect of ignoring the time-series error, it is instructive to examine the ratio between (36) and (37) as $n \to \infty$ with τ fixed. Simple calculations indicate that the ratio (36)/(37) converges to zero. This implies that for n sufficiently large, the erroneous standard error will be substantially smaller than the correct standard error. As a consequence, the confidence intervals based on the erroneous standard error will have smaller coverage probability as n becomes larger, which explains the first pattern in our Monte Carlo exercise. This problem is less severe if τ is sufficiently large that we can ignore the second term in (37), since in this case the only difference between (36) and (37) is given by the $\hat{A}_{f,\rho} \hat{A}_{g,\rho}^{-1} \hat{A}_{g,\theta}$ inside the first term of (37), which explains the second pattern in the Monte Carlo exercise.

We now describe the estimation of the risk aversion parameter using only cross-sectional data. As discussed in Section 5, the parameter μ requires both cross-section and time-series variation to be consistently estimated. If the econometrician uses only cross-sectional data, the estimated μ will be biased. In Table 3 we report the mean and median estimated μ and the corresponding bias only for the three time-series, since the results are nearly identical across cross-sections. The numbers indicate that the bias is positive, extremely large, and similar for all time-series whether one considers the mean or the median. The parameter μ is estimated to be between six and seven times the size of the true parameter and the bias to be between five and six times the true value. A bias of this magnitude can have significant consequences if the estimated parameter is used to answer policy questions, with answers that can be considerably different from the ones that should be obtained.

In Tables 4 and 5, we also report the effect of changing the persistence of the aggregate shock by increasing ρ from 0.75 to 0.9 and by reducing it from 0.75 to 0.5 for the the specification with 10,000 people and 100 periods. The effect is small. When we use our proposed method the estimated coefficients are close to the true values. But if one ignores the aggregate shocks the bias is large and positive.

Our Monte Carlo results indicate that ignoring aggregate shocks that affect the data can have large effects on inference, on the estimation of important parameters, such as the coefficient of risk aversion, and on the policy evaluations which are based on them. Our results also indicate that the inference and estimation method we propose performs well. Given that it is relatively straightforward to use, it is an easy solution for dealing with the presence of aggregate shocks.

8 Summary

Using a general econometric framework and three examples we shown that generally, when aggregate shocks are present, model parameters cannot be identified using cross-sectional variation alone. Identification of those parameters requires the combination of cross-sectional and time-series data. When those two data sources are jointly used, there are no available formulas for the computation of test statistics and confidence intervals. We provide new easy-to-use formulas that account for the interaction between those data sources. Our results are expected to be helpful for the econometric analysis of rational expectations models involving individual decision making as well as general equilibrium models.

References

[1] Ackerberg, D.A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," Econometrica 83, 2411–2451.

[2] Aldous, D.J. and G.K. Eagleson (1978), On Mixing and Stability of Limit Theorems, *The Annals of Probability*, 6, 325-331.

[3] Altug, S., and Miller, R. (1990): "Household Choices in Equilibrium," Econometrica, 58(3), 543-570.

[4] Andrews, D.W.K. (2005): "Cross-Section Regression with Common Shocks," Econometrica 73, pp. 1551-1585.

[5] Arellano, M., R. Blundell, and S. Bonhomme (2014): "Household Earnings and Consumption: A Nonlinear Framework," unpublished working paper.

[6] Bajari, Patrick, Benkard, C. Lanier, and Levin, Jonathan (2007): "Estimating Dynamic Models of Imperfect Competition," Econometrica 75, 5, pp. 1331-1370.

[7] Barsky, Robert B., and Lutz Kilian (2004): "Oil and the Macroeconomy since the 1970s," The Journal of Economic Perspectives, vol. 18, no. 4, pp. 115–134.

[8] Chamberlain, G. (1984): "Panel Data," in Handbook of Econometrics, eds. by Z. Griliches and M. Intriligator. Amsterdam: North Holland, pp. 1247-1318.

[9] Cho, I, T.J. Sargent, and N. Williams (2002): "Escaping Nash Inflation," Review of Economic Studies 69, pp. 1–40.

[10] Davidson, C., and Woodbury, S. (1993): "The Displacement Effect of Reemployment Bonus Programs," Journal of Labor Economics, 11(4), 575-605.

[11] Dix-Carneiro, Rafael (2014): "Trade Liberalization and Labor Market Dynamics," Econometrica, 82: 825-885.

[12] Eckstein, Zvi, and Osnat Lifshitz. "Dynamic Female Labor Supply."Econometrica, vol. 79, no. 6, 2011, pp. 1675–1726.

[13] Ferreyra, Maria Marta (2007): "Estimating the Effects of Private School Vouchers in Multi-district Economies," American Economic Review, 97(3), 789-817.

[14] Gagliardini, P., and C. Gourieroux (2011): "Efficiency in Large Dynamic Panel Models with Common Factor," unpublished working paper.

[15] Gandhi, A., Navarro, S., and Rivers, D. (2011): "On the Identification of Production Functions: How Heterogeneous is Productivity?" unpublished working paper.

[16] Gemici, A., and M. Wiswall (2014): "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors," International Economic Review 55, 23–56.

[17] Gillingham, K., F. Iskhakov, A. Munk-Nielsen, J. Rust, and B. Schjerning (2015): "A Dynamic Model of Vehicle Ownership, Type Choice, and Usage," unpublished working paper.

[18] Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," Econometrica 70, pp. 1639–57.

[19] Hahn, J., Kuersteiner, G. and Mazzocco, M (2016): "Central Limit Theory for Combined Cross-Section and Time Series,"Working Paper.

[20] Hahn, J., and W.K. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Econometrica 72, pp. 1295–1319.

[21] Hansen, L.P., T.J. Sargent, and T. Tallarini (1999): "Robust Permanent Income and Pricing," Review of Economic Studies 66, pp. 873–907.

[22] Heckman, J.J., L. Lochner, and C. Taber (1998), "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," Review of Economic Dynamics 1, pp. 1-58.

[23] Heckman, J.J., and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," Journal of Political Economy, 93, pp. 1077-1125.

[24] Keane, M., and Wolpin, K. (1997): "The Career Decisions of Young Men." Journal of Political Economy, 105(3), 473–52.

[25] Kydland, F.E., and E.C. Prescott (1982): "Time to Build and Aggregate Fluctuations," Econometrica 50, pp. 1345–1370.

[26] Kuersteiner, G.M., and I.R. Prucha (2013): "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," Journal of Econometrics 174, pp. 107-126.

[27] Kuersteiner, G.M and I.R. Prucha (2015): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," CESifo Working Paper No. 5445.

[28] Lee, D. (2005): "An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice," International Economic Review 46, pp. 1-34.

[29] Lee, D., and K.I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of Service Sector," Econometrica 47, pp. 1-46.

[30] Lee, D., and K.I. Wolpin (2010): "Accounting for Wage and Employment Changes in the U.S. from 1968-2000: A Dynamic Model of Labor Market Equilibrium," Journal of Econometrics 156, pp. 68–85.

[31] Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables", Review of Economic Studies 70, pp. 317–341.

[32] Lise, J., S. Seitz, and J. Smith (2004): "Equilibrium Policy Experiments and the Evaluation of Social Programs," NBER Working Paper Series No. 10283.

[33] Macurdy, T.E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", Journal of Econometrics 18, pp. 83–114.

[34] Mehta, N. (2017): "Competition In Public School Districts: Charter School Entry, Student Sorting, And School Input Determination," International Economic Review, 58: 1089-1116.

[35] Murphy, K. M. and R. H. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," Journal of Business and Economic Statistics 3, pp. 370 – 379.

[36] Olley, G.S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," Econometrica 64, pp. 1263 – 1297.

[37] Renyi, A (1963): "On stable sequences of events," *Sankya Ser. A*, 25, 293-302.

[38] Runkle, D.E. (1991): "Liquidity Constraints and the Permanent-Income Hypothesis: Evidence from Panel Data," Journal of Monetary Economics 27, pp. 73–98.

[39] Shea, J. (1995):"Union Contracts and the Life-Cycle/Permanent-Income Hypothesis," American Economic Review 85, pp. 186–200.

[40] Shephard, A. (2017): "Equilibrium Search And Tax Credit Reform,"International Economic Review, 58: 1047-1088.

Table 1: Monte Carlo Results, Parameter Estimates For Correct Model

True Parameter	Estimate	Cov. Prob.	Cov. Prob. No Agg. Uncert.
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 25			
Log Risk Aversion Mean: $\mu = 0.2$	**0.161**	**0.903**	**0.659**
Aggregate Shock Persistence: $\rho = 0.75$	0.698	0.872	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.514	0.833	-
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 50			
Log Risk Aversion Mean: $\mu = 0.2$	**0.172**	**0.919**	**0.742**
Aggregate Shock Persistence: $\rho = 0.75$	0.722	0.880	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.502	0.870	-
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.173**	**0.936**	**0.818**
Aggregate Shock Persistence: $\rho = 0.75$	0.736	0.891	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.498	0.888	-
Infeasible estimator of Log Risk Aversion Mean, Cross-section of 2,500: 0.1997			
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 25			
Log Risk Aversion Mean: $\mu = 0.2$	**0.175**	**0.902**	**0.587**
Aggregate Shock Persistence: $\rho = 0.75$	0.698	0.868	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.508	0.830	-
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 50			
Log Risk Aversion Mean: $\mu = 0.2$	**0.183**	**0.916**	**0.670**
Aggregate Shock Persistence: $\rho = 0.75$	0.722	0.876	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.497	0.873	-
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.173**	**0.932**	**0.758**
Aggregate Shock Persistence: $\rho = 0.75$	0.736	0.889	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.492	0.894	-
Infeasible estimator of Log Risk Aversion Mean, Cross-section of 5,000: 0.1997			
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 25			
Log Risk Aversion Mean: $\mu = 0.2$	**0.173**	**0.890**	**0.515**
Aggregate Shock Persistence: $\rho = 0.75$	0.698	0.868	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.503	0.826	-
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 50			
Log Risk Aversion Mean: $\mu = 0.2$	**0.184**	**0.913**	**0.596**
Aggregate Shock Persistence: $\rho = 0.75$	0.722	0.872	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.492	0.865	-
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.184**	**0.928**	**0.694**
Aggregate Shock Persistence: $\rho = 0.75$	0.737	0.890	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.487	0.887	-
Infeasible estimator of Log Risk Aversion Mean, Cross-section of 10,000: 0.200			

Notes: This table reports the Monte Carlo results for the correct model obtained using our proposed estimation method. They are derived by simulating the general equilibrium model 5000 times. The second column reports the average estimated parameter, where the average is computed over the 5000 simulations. Column 3 reports the coverage probability of a confidence interval with 90% nominal coverage probability.

Table 2: Monte Carlo Results, Parameter Estimates For Correct Model, Cont.

True Parameter	Estimate	Cov. Prob.	Cov. Prob. No Agg. Uncert.
Cross-sectional Sample Size: 50,000, Time-series Sample Size: 25			
Log Risk Aversion Mean: $\mu = 0.2$	**0.176**	**0.876**	**0.296**
Aggregate Shock Persistence: $\rho = 0.75$	0.698	0.868	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.502	0.820	-
Cross-sectional Sample Size: 50,000, Time-series Sample Size: 50			
Log Risk Aversion Mean: $\mu = 0.2$	**0.187**	**0.897**	**0.377**
Aggregate Shock Persistence: $\rho = 0.75$	0.722	0.870	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.491	0.854	-
Cross-sectional Sample Size: 50,000, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.189**	**0.911**	**0.469**
Aggregate Shock Persistence: $\rho = 0.75$	0.736	0.882	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.487	0.884	-
Infeasible estimator of Log Risk Aversion Mean, Cross-section of 5,000:			0.1994

Notes: See table 1.

Table 3: Monte Carlo Results, Risk Aversion Estimates For Misspecified Model

	Mean		Median	
True Parameter	Estimate	Bias	Estimate	Bias
Cross-sectional Sample Size: 2,500				
Log Risk Aversion Mean: $\mu = 0.2$	**1.224**	**1.024**	**1.400**	**1.200**
Cross-sectional Sample Size: 5,000				
Log Risk Aversion Mean: $\mu = 0.2$	**1.224**	**1.024**	**1.405**	**1.205**
Cross-sectional Sample Size: 10,000				
Log Risk Aversion Mean: $\mu = 0.2$	**1.227**	**1.027**	**1.416**	**1.216**
Cross-sectional Sample Size: 50,000				
Log Risk Aversion Mean: $\mu = 0.2$	**1.229**	**1.029**	**1.416**	**1.216**

Notes: This table reports the Monte Carlo results for the misspecified model obtained using only cross-sectional variation. They are derived by simulating the general equilibrium model 5000 times. The second column reports the average estimated parameter, where the average is computed over the 5000 simulations. Column 3 reports the estimation bias, which is computed as the difference between the estimated and true parameter.

Table 4: Monte Carlo Results, Parameter Estimates For Correct Model, Different ρ's

True Parameter	Estimate	Cov. Prob.	Cov. Prob. No Agg. Uncert.
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.217**	**0.911**	**0.726**
Aggregate Shock Persistence: $\rho = 0.9$	0.884	0.899	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.493	0.883	-
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 100			
Log Risk Aversion Mean: $\mu = 0.2$	**0.180**	**0.929**	**0.654**
Aggregate Shock Persistence: $\rho = 0.5$	0.492	0.884	-
Variance of Aggregate Shock: $\omega^2 = 0.48$	0.489	0.886	-

See notes at Table 1.

Table 5: Monte Carlo Results, Risk Aversion Estimates For Misspecified Model, Different ρ's

	Mean		Median	
True Parameter	Estimate	Bias	Estimate	Bias
Cross-sectional Sample Size: 10,000, $\rho = 0.9$				
Log Risk Aversion Mean: $\mu = 0.2$	**1.192**	**0.992**	**1.342**	**1.142**
Cross-sectional Sample Size: 10,000, $\rho = 0.5$				
Log Risk Aversion Mean: $\mu = 0.2$	**1.208**	**1.008**	**1.401**	**1.201**

See notes at Table 3.

On line Appendix

A Discussion for Section 3

A.1 Proof of (2)

The maximization problem is equivalent to

$$\max_{\alpha} \; -e^{-\delta(\alpha(1+r)+(1-\alpha))} E\left[e^{-\delta(1-\alpha)u_{i,t}}\right].$$

Since $-\delta\,(1-\alpha)\,u_{i,t} \sim N\left(-\delta\,(1-\alpha)\,\mu, \delta^2\,(1-\alpha)^2\,\sigma^2\right)$, we have

$$E\left[e^{-\delta(1-\alpha)u_{i,t}}\right] = e^{-\delta(1-\alpha)\mu+\frac{\delta^2(1-\alpha)^2\sigma^2}{2}},$$

and the maximization problem can be rewritten as follows:

$$\max_{\alpha} \; -e^{-\delta\left(\alpha(1+r)+(1-\alpha)(1+\mu)-\frac{\delta(1-\alpha)^2\sigma^2}{2}\right)}.$$

Taking the first order condition, we have,

$$0 = -\delta\left(r-\mu+\sigma^2\delta-\alpha\sigma^2\delta\right),$$

from which we obtain the solution

$$\alpha = \frac{1}{\sigma^2\delta}\left(r-\mu+\sigma^2\delta\right).$$

A.2 Euler Equation and Cross Section

Our model in Section 3 is a stylized version of many models considered in a large literature interested in estimating the parameter δ using cross-sectional variation. Estimators are often based on moment conditions derived from first order conditions (FOC) related to optimal investment and consumption decisions. We illustrate the problems facing such estimators.

Assume a researcher has a cross-section of observations for individual consumption and returns $c_{i,t}$ and $u_{i,t}$. The population FOC of our model[32] takes the simple form $E\left[e^{-\delta c_{i,t}}\left(r-u_{i,t}\right)\right]=0$. A just-identified moment based estimator for δ solves the sample analog $n^{-1}\sum_{i=1}^{n}e^{-\hat{\delta}c_{i,t}}\left(r-u_{i,t}\right)=0$. It turns out that the probability limit of $\hat{\delta}$ is equal to $\left(\nu_t-r\right)/\left(\left(1-\alpha\right)\sigma_\epsilon^2\right)$, i.e., $\hat{\delta}$ is inconsistent.

We now compare the population FOC a rational agent uses to form their optimal portfolio with the empirical FOC an econometrician using cross-sectional data observes:

$$n^{-1}\sum_{i=1}^{n}e^{-\delta c_{i,t}}\left(r-u_{i,t}\right)=0.$$

Noting that $u_{i,t}=\nu_t+\epsilon_{i,t}$ and substituting into the budget constraint

$$c_{i,t}=1+\alpha r+\left(1-\alpha\right)u_{i,t}=1+\alpha r+\left(1-\alpha\right)\nu_t+\left(1-\alpha\right)\epsilon_{i,t},$$

we have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}e^{-\delta c_{i,t}}\left(r-u_{i,t}\right) &= n^{-1}\sum_{i=1}^{n}e^{-\delta\left(1+\alpha r+\left(1-\alpha\right)\nu_t\right)-\delta\left(1-\alpha\right)\epsilon_{i,t}}\left(r-\nu_t-\epsilon_{i,t}\right) \\ &= e^{-\delta\left(1+\alpha r+\left(1-\alpha\right)\nu_t\right)}\left(\left(r-\nu_t\right)n^{-1}\sum_{i=1}^{n}e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}-n^{-1}\sum_{i=1}^{n}e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}\right). \end{aligned} \tag{38}$$

Under suitable regularity conditions including independence of $\epsilon_{i,t}$ in the cross-section it follows that

$$n^{-1}\sum_{i=1}^{n}e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}=E\left[e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\right]+o_p\left(1\right)=e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}+o_p\left(1\right) \tag{39}$$

and

$$n^{-1}\sum_{i=1}^{n}e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}=E\left[e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}\right]+o_p\left(1\right)=-\delta\left(1-\alpha\right)\sigma_\epsilon^2e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}+o_p\left(1\right). \tag{40}$$

Taking limits as $n\to\infty$ in (38) and substituting (39) and (40) then shows that the method of moments estimator based on the empirical FOC asymptotically solves

$$\left(\left(r-\nu_t\right)+\delta\left(1-\alpha\right)\sigma_\epsilon^2\right)e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}=0. \tag{41}$$

[32] We assume $\delta\neq0$ and rescale the equation by $-\delta^{-1}$.

Solving for δ we obtain

$$\operatorname{plim}\hat{\delta} = \frac{\nu_t - r}{(1-\alpha)\,\sigma_\epsilon^2}.$$

This estimate is inconsistent because the cross-sectional data set lacks cross sectional ergodicity, or in other words does not contain the same information about aggregate risk as is used by rational agents. Therefore, the empirical version of the FOC is unable to properly account for aggregate risk and return characterizing the risky asset. The estimator based on the FOC takes the form of an implicit solution to an empirical moment equation, which obscures the effects of cross-sectional non-ergodicity. A more illuminative approach uses our modelling strategy in Section 2.

On the other hand, it is easily shown using properties of the Gaussian moment generating function that the population FOC is proportional to

$$E\left[e^{-\delta(1-\alpha)u_{i,t}}\left(r-u_{i,t}\right)\right] = \left(r-\mu+\delta\,(1-\alpha)\,\sigma^2\right)e^{-\delta(1-\alpha)\mu+\frac{\delta^2(1-\alpha)^2\sigma^2}{2}} = 0. \tag{42}$$

The main difference between (39) and (40) lies in the fact that σ_v^2 is estimated to be 0 in the sample and that $\nu_t \neq \mu$ in general. Note that (42) implies that consistency may be achieved with a large number of repeated cross sections, or a panel data set with a long time series dimension. However, this raises other issues discussed in Section C.

B Details of Section 5

B.1 Proof of (25)

In the proof we will drop the i subscripts for notational purposes. The individual will choose education F if

$$E\left[V_{t+1}\left(F,\nu_{t+1},\varepsilon_{t+1}^F\right)\middle|\,\nu_t\right] \geq E\left[V_{t+1}\left(R,\nu_{t+1},\varepsilon_{t+1}^R\right)\middle|\,\nu_t\right].$$

Using (52) and (53) later in Section B.2, we write

$$V_{t+1}\left(F,\nu_{t+1},\varepsilon_{t+1}^F\right) = \frac{\left[\left(\left(\frac{n_{t+1}^F\sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1}\sigma T\right)^{\sigma}\left((1-\sigma)\,T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}\exp\left(\sigma\,(1/\alpha_1)\,(1-\gamma)\,\varepsilon_{t+1}^F\right),$$

and

$$V_{t+1}\left(R,\nu_{t+1},\varepsilon^{R}_{t+1}\right)=\frac{\left[\left(\left(\frac{n^{R}_{t+1}\sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1}\sigma T\right)^{\sigma}\left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon^{R}_{t+1}\right)$$
$$\times\left(\nu_{t+1}^{-\sigma(1/\alpha_1)(1-\gamma)}\right).$$

It follows that education F is chosen if and only if

$$\left(n^{F}_{t+1}\right)^{\sigma(1-\gamma)/\alpha_1}\geq\left(n^{R}_{t+1}\right)^{\sigma(1-\gamma)/\alpha_1}\times\frac{E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon^{R}_{t+1}\right)\right]E_t\left[\nu_{t+1}^{-\sigma(1/\alpha_1)(1-\gamma)}\right]}{E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon^{F}_{t+1}\right)\right]}. \tag{43}$$

Recall that $E\left[\exp\left(\varepsilon^{S}_{t}\right)\right]=1$ for $S=F,R$. It follows that $\varepsilon^{F}_{t+1}\sim N\left(-\frac{\sigma^2_F}{2},\sigma^2_F\right)$, and $\varepsilon^{R}_{t+1}\sim N\left(-\frac{\sigma^2_R}{2},\sigma^2_R\right)$, and as a consequence,

$$E\left[\exp\left(\frac{\sigma\left(1-\gamma\right)}{\alpha_1}\varepsilon^{F}_{t+1}\right)\right]=\exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\sigma^2_F}{2}+\frac{\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\right)^2\sigma^2_F}{2}\right), \tag{44}$$
$$E\left[\exp\left(\frac{\sigma\left(1-\gamma\right)}{\alpha_1}\varepsilon^{R}_{t+1}\right)\right]=\exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\sigma^2_R}{2}+\frac{\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\right)^2\sigma^2_R}{2}\right). \tag{45}$$

Also, because $\log\nu_{t+1}=\rho\log\nu_t+\eta_t$, or $\nu_{t+1}=\nu_t^{\rho}\exp\left(\eta_t\right)$, we can write

$$E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right]=E_\eta\left[\left(\nu_t^{\rho}\exp\left(\eta_t\right)\right)^{-\sigma(1-\gamma)(1/\alpha_1)}\right]=\nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)}E\left[\exp\left(-\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\eta_t\right)\right].$$

where $E_\eta\left[\cdot\right]$ denotes the integral with respect to η_t alone. The assumption that $\eta_t\sim N\left(0,\omega^2\right)$ allows us to write

$$E\left[\exp\left(-\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\eta_t\right)\right]=\exp\left(\frac{\left(\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\right)^2}{2}\omega^2\right)$$

recognizing that the expectation on the left is nothing but the moment generating function of

$N(0,\omega^2)$ evaluated at $-\sigma(1-\gamma)(1/\alpha_1)$. Therefore, we have

$$E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right] = \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{(\sigma(1-\gamma)(1/\alpha_1))^2}{2}\omega^2\right). \tag{46}$$

Combining (44), (45), and (46), we obtain

$$\begin{aligned}
&\frac{E\left[\exp\left(\sigma(1/\alpha_1)(1-\gamma)\varepsilon_{t+1}^R\right)\right] E_t\left[\nu_{t+1}^{-\sigma(1/\alpha_1)(1-\gamma)}\right]}{E\left[\exp\left(\sigma(1/\alpha_1)(1-\gamma)\varepsilon_{t+1}^F\right)\right]} \\
&= \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{(\sigma(1-\gamma)(1/\alpha_1))^2}{2}\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)\right) \\
&\times \exp\left(-\frac{\sigma(1/\alpha_1)(1-\gamma)(\sigma_R^2 - \sigma_F^2)}{2}\right).
\end{aligned}$$

As a consequence, (43) is equivalent to

$$\begin{aligned}
\left(n_{t+1}^F\right)^{\sigma(1-\gamma)/\alpha_1} \geq \left(n_{t+1}^R\right)^{\sigma(1-\gamma)/\alpha_1} \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{(\sigma(1-\gamma)(1/\alpha_1))^2}{2}\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)\right) \\
\times \exp\left(-\frac{\sigma(1/\alpha_1)(1-\gamma)(\sigma_R^2 - \sigma_F^2)}{2}\right)
\end{aligned} \tag{47}$$

when $1-\gamma > 0$, and

$$\begin{aligned}
\left(n_{t+1}^F\right)^{\sigma(1-\gamma)/\alpha_1} \leq \left(n_{t+1}^R\right)^{\sigma(1-\gamma)/\alpha_1} \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{(\sigma(1-\gamma)(1/\alpha_1))^2}{2}\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)\right) \\
\times \exp\left(-\frac{\sigma(1/\alpha_1)(1-\gamma)(\sigma_R^2 - \sigma_F^2)}{2}\right)
\end{aligned} \tag{48}$$

when $1-\gamma < 0$.

Consider first the case $1-\gamma > 0$. Taking logs of (47), we obtain

$$\begin{aligned}
\frac{\sigma(1-\gamma)}{\alpha_1}\log n_{t+1}^F \geq \frac{\sigma(1-\gamma)}{\alpha_1}\log n_{t+1}^R - \rho\frac{\sigma(1-\gamma)}{\alpha_1}\log\nu_t \\
+ \frac{(\sigma(1-\gamma))^2}{2\alpha_1^2}\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right) - \frac{\sigma(1/\alpha_1)(1-\gamma)(\sigma_R^2 - \sigma_F^2)}{2}.
\end{aligned}$$

Dividing by σ and multiplying by $\alpha_1 < 0$, we conclude that the decision is equivalent to

$$(1-\gamma)\left(\log\frac{n_{t+1}^F}{n_{t+1}^R}+\frac{(\sigma_R^2-\sigma_F^2)}{2}+\rho\log\nu_t\right)\leq\frac{\sigma(1-\gamma)^2(\sigma_R^2-\sigma_F^2+\omega^2)}{2\alpha_1}.$$

Dividing by $\sigma(1-\gamma)(\sigma_R^2-\sigma_F^2+\omega^2)>0$, we obtain

$$\frac{\log\frac{n_{t+1}^F}{n_{t+1}^R}+\frac{\sigma_R^2-\sigma_F^2}{2}+\rho\log\nu_t}{\sigma(\sigma_R^2-\sigma_F^2+\omega^2)}\leq\frac{1-\gamma}{2\alpha_1}.$$

Multiplying by $2\alpha_1 < 0$, we obtain

$$\frac{\log\frac{n_{t+1}^F}{n_{t+1}^R}+\frac{\sigma_R^2-\sigma_F^2}{2}+\rho\log\nu_t}{\frac{\sigma\left(\sigma_R^2-\sigma_F^2+\omega^2\right)}{2\alpha_1}}\geq 1-\gamma$$

or

$$\gamma\geq 1-\frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right)+\frac{\sigma_R^2-\sigma_F^2}{2}+\rho\log\nu_t}{\frac{\sigma\left(\sigma_R^2-\sigma_F^2+\omega^2\right)}{2\alpha_1}},$$

which proves inequality (25) for the $1-\gamma>0$ case.

Consider now the case $1-\gamma<0$. Taking logs of (48), we obtain

$$\begin{aligned}\frac{\sigma(1-\gamma)}{\alpha_1}\log n_{t+1}^F\leq\frac{\sigma(1-\gamma)}{\alpha_1}\log n_{t+1}^R-\rho\frac{\sigma(1-\gamma)}{\alpha_1}\log\nu_t\\+\frac{(\sigma(1-\gamma))^2}{2\alpha_1^2}\left(\sigma_R^2-\sigma_F^2+\omega^2\right)-\frac{\sigma(1/\alpha_1)(1-\gamma)(\sigma_R^2-\sigma_F^2)}{2}.\end{aligned}$$

Dividing by σ and multiplying by $\alpha_1 < 0$, we conclude that the decision is equivalent to

$$(1-\gamma)\left(\log\frac{n_{t+1}^F}{n_{t+1}^R}+\frac{\sigma_R^2-\sigma_F^2}{2}+\rho\log\nu_t\right)\geq\frac{\sigma(1-\gamma)^2(\sigma_R^2-\sigma_F^2+\omega^2)}{2\alpha_1}.$$

Dividing by by $\sigma(1-\gamma)(\sigma_R^2-\sigma_F^2+\omega^2)<0$, we obtain

$$\frac{\log\frac{n_{t+1}^F}{n_{t+1}^R}+\frac{\sigma_R^2-\sigma_F^2}{2}+\rho\log\nu_t}{\sigma(\sigma_R^2-\sigma_F^2+\omega^2)}\leq\frac{(1-\gamma)}{2\alpha_1}.$$

Multiplying by $2\alpha_1 < 0$, we obtain

$$\frac{\log \frac{n^F_{t+1}}{n^R_{t+1}} + \frac{\sigma_R^2 - \sigma_F^2}{2} + \rho \log \nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}} \geq 1 - \gamma$$

or

$$\gamma \geq 1 - \frac{\log \frac{n^F_{t+1}}{n^R_{t+1}} + \frac{\sigma_R^2 - \sigma_F^2}{2} + \rho \log \nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}},$$

which proves inequality (25) for the $1 - \gamma < 0$ case as well.

B.2 Proof of (23) and (24)

Note that individual heterogeneity is completely summarized by the vector $\chi_t \equiv \left(\varepsilon^F_t, \varepsilon^R_t, \gamma\right)$. This means that the labor supply for each type χ of workers can be written $h^F_t(\chi)$ and $h^R_t(\chi)$. We assume that the measure of individuals such that $\left(\varepsilon^F_t, \varepsilon^R_t, \gamma\right) \in A$ for some $A \subset R^3$ is given by $N_t \int_A G(d\chi)$, where G is a joint CDF. For simplicity, we assume that G is such that the first and second components are independent of each other. Recall that we also assume that $\int \exp(\varepsilon_t) G(d\chi) = 1$.

We can rewrite (22) as follows:

$$\begin{aligned} & E\left[\left.\frac{\left[\left(\sigma w^F_{t+1}(\nu_{t+1}) \exp\left(\varepsilon^F_{t+1}\right) T\right)^{\sigma} \left((1-\sigma) T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}\right| \nu_t\right] \\ & \geq E\left[\left.\frac{\left[\left(\sigma w^R_{t+1}(\nu_{t+1}) \exp\left(\varepsilon^R_{t+1}\right) T\right)^{\sigma} \left((1-\sigma) T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}\right| \nu_t\right]. \end{aligned} \tag{49}$$

As a consequence, education F is chosen if

$$\psi(\gamma, \nu_t) \equiv E\left[\left.\frac{\left[\left(w^F_{t+1}(\nu_{t+1}) \exp\left(\varepsilon^F_{t+1}\right)\right)\right]^{\sigma(1-\gamma)}}{1-\gamma}\right| \nu_t\right] - E\left[\left.\frac{\left[w^R_{t+1}(\nu_{t+1}) \exp\left(\varepsilon^R_{t+1}\right)\right]^{\sigma(1-\gamma)}}{1-\gamma}\right| \nu_t\right] \geq 0. \tag{50}$$

Specifically, an individual chooses F if $\psi(\gamma, \nu_t) > 0$. We can now introduce the equilibrium condition for education F. It takes the following form:

$$\begin{aligned} H_{t+1}^{D,F} &= N_{t+1} \int_{E=F} h_{t+1}^{F}(\chi)\, G(d\chi) = N_{t+1}\sigma\mathcal{T} \int_{\psi(\gamma,\nu_t)\geq 0} \exp\left(\varepsilon_{t+1}^{F}\right) G(d\chi) \\ H_{t+1}^{D,R} &= N_{t+1} \int_{E=R} h_{t+1}^{F}(\chi)\, G(d\chi) = N_{t+1}\sigma\mathcal{T} \int_{\psi(\gamma,\nu_t)< 0} \exp\left(\varepsilon_{t+1}^{R}\right) G(d\chi). \end{aligned}$$

Using independence between γ and ε as well as $\int \exp\left(\varepsilon_t^F\right) G(d\chi) = 1$, we can write

$$\begin{aligned} \int_{\psi(\gamma,\nu_t)\geq 0} \exp\left(\varepsilon_{t+1}^{F}\right) G(d\chi) &= \left(\int_{\psi(\gamma,\nu_t)\geq 0} G(d\chi)\right)\left(\int \exp\left(\varepsilon_{t+1}^{F}\right) G(d\chi)\right) \\ &= \int_{\psi(\gamma,\nu_t)\geq 0} G(d\chi) \\ &= \text{Fraction of workers in Sector } F, \end{aligned} \tag{51}$$

so we can write $H_t^{D,F} = n_t^F \sigma T$, where n^F is the measure of individuals that chose education F. Taking logs, we have:

$$\log H_t^{D,F} = \log n_t^F + \log\sigma + \log T,$$

Substituting for $H_t^{D,F}$, we obtain the following equilibrium condition:

$$\alpha_0 + \alpha_1 \log w_t^F = \log n_t^F + \log\sigma + \log T,$$

Solving for $\log w_t^F$, we have the log equilibrium wage:

$$\left(z_t^F \equiv\right) \quad \log w_t^F = \frac{\log n_t^F + \log\sigma + \log T - \alpha_0}{\alpha_1}. \tag{52}$$

This wage is for the unit of effective labor. Because the worker i provides $\sigma\exp(\varepsilon_t) T$ of effective labor, his recorded earning is $\sigma\exp(\varepsilon_t) T \exp\left(\frac{\log n_t^F + \log\sigma + \log T - \alpha_0}{\alpha_1}\right)$. Because the individual works for σT hours, his wage for the labor is $\exp(\varepsilon_t)\exp\left(\frac{\log n_t^F + \log\sigma + \log T - \alpha_0}{\alpha_1}\right)$; we will assume that the cross section "error" consist of n i.i.d. copies of ε_t, i.e.,the observed log equilibrium individual wage follows:

$$\log w_{it}^F = \frac{\log n_t^F + \log\sigma + \log T - \alpha_0}{\alpha_1} + \varepsilon_{it}^F.$$

Because of the normalization $E\left[\exp\left(\varepsilon_{it}^R\right)\right] = 1$, the second equality in (51) also applies to the R sector, and as a consequence, the equilibrium condition for education R has the following form:

$$H_t^{D,R} = n_t^R \sigma T,$$

where n^R is the measure of individuals that chose education R. Substituting for $H_t^{D,R}$ and solving for $\log w_t^R$, we obtain the following equilibrium wage for R:

$$\left(z_t^R \equiv\right) \quad \log w_t^R = \frac{\log n_t^R + \log \sigma + \log T - \alpha_0 - \log \nu_t}{\alpha_1}. \tag{53}$$

By the same reasoning, the observed log equilibrium wage would look like

$$\log w_{it}^R = \frac{\log n_t^R + \log \sigma + \log T - \alpha_0 - \log \nu_t}{\alpha_1} + \varepsilon_{it}^R.$$

C Long Panels?

Our proposal requires access to two data sets, a cross-section (or short panel) and a long time series of aggregate variables. One may wonder whether we may obtain an estimator with similar properties by exploiting panel data sets in which the time series dimension of the panel data is large enough.

One obvious advantage of combining two sources of data is that time series data may contain variables that are unavailable in typical panel data sets. For example the inflation rate potentially provides more information about aggregate shocks than is available in panel data. We argue with a toy model that even without access to such variables, the estimator based on the two data sets is expected to be more precise, which suggests that the advantage of data combination goes beyond availability of more observable variables.

Consider the alternative method based on one long panel data set, in which both n and $\mathbb{T}$ go to infinity. Since the number of aggregate shocks ν_t increases as the time-series dimension $\mathbb{T}$ grows, we expect that the long panel analysis can be executed with tedious yet straightforward arguments by modifying ideas in Hahn and Kuersteiner (2002), Hahn and Newey (2004) and Gagliardini and Gourieroux (2011), among others.

We will now illustrate a potential problem with the long panel approach with a simple artificial example. Suppose that the econometrician is interested in the estimation of a parameter γ that characterizes the following system of linear equations:

$$\begin{aligned} q_{i,t} &= x_{i,t}\frac{\gamma}{\omega} + \nu_t + \varepsilon_{i,t} \qquad i = 1, \ldots, n;\ t = 1, \ldots, \mathbb{T}, \\ \nu_t &= \omega \nu_{t-1} + u_t. \end{aligned}$$

The variables $q_{i,t}$ and $x_{i,t}$ are observed and it is assumed that $x_{i,t}$ is strictly exogenous in the sense that it is independent of the error term $\varepsilon_{i,t}$, including all leads and lags. For simplicity, we also assume that u_t and $\varepsilon_{i,t}$ are normally distributed with zero mean and that $\varepsilon_{i,t}$ is i.i.d. across both i and t. We will denote by δ the ratio $\gamma/\,\omega$.

In order to estimate γ based on the panel data $\{(q_{i,t}, x_{i,t}), i = 1, \ldots, n;\ t = 1, \ldots, \mathbb{T}\}$, we can adopt a simple two-step estimator of γ. In a first step, the parameter δ and the aggregate shocks ν_t are estimated using an Ordinary Least Square (OLS) regression of $q_{i,t}$ on $x_{i,t}$ and time dummies. In the second step, the time-series parameter ω is estimated by regressing $\widehat{\nu}_t$ on $\widehat{\nu}_{t-1}$, where $\widehat{\nu}_t$, $t = 1, \ldots, \mathbb{T}$, are the aggregate shocks estimated in the first step using the time dummies. An estimator of γ can then be obtained as $\widehat{\delta}\widehat{\omega}$.

The following remarks are useful to understand the properties of the estimator $\widehat{\gamma} = \widehat{\delta}\widehat{\omega}$. First, even if ν_t were observed, for $\widehat{\omega}$ to be a consistent estimator of ω we would need $\mathbb{T}$ to go to infinity, under which assumption we have $\widehat{\omega} = \omega + O_p\left(\mathbb{T}^{-1/2}\right)$. This implies that it is theoretically necessary to assume that our data source is a "long" panel, i.e., $\mathbb{T} \to \infty$. Similarly, $\hat{\nu}_t$ is a consistent estimator of ν_t only if n goes to infinity. As a consequence, we have $\hat{\nu}_t = \nu_t + O_p\left(n^{-1/2}\right)$. This implies that it is in general theoretically necessary to assume that $n \to \infty$.[33] Moreover, if n and $\mathbb{T}$ both go to infinity, $\widehat{\delta}$ is a consistent estimator of δ and $\widehat{\delta} = \delta + O_p\left(n^{-1/2}\mathbb{T}^{-1/2}\right)$. All this implies that

$$\widehat{\gamma} = \widehat{\delta}\widehat{\omega} = \left(\delta + O_p\left(\frac{1}{\sqrt{n\mathbb{T}}}\right)\right)\left(\omega + O_p\left(\frac{1}{\sqrt{\mathbb{T}}}\right)\right) = \delta\omega + O_p\left(\frac{1}{\sqrt{\mathbb{T}}}\right) = \gamma + O_p\left(\frac{1}{\sqrt{\mathbb{T}}}\right).$$

The $O_p\left(n^{-1/2}\mathbb{T}^{-1/2}\right)$ estimation noise of $\widehat{\delta}$, which is dominated by the $O_p\left(\mathbb{T}^{-1/2}\right)$ error from

[33] For $\widehat{\omega}$ to have the same distribution as if ν_t were observed, we need n to go to infinity faster than T or equivalently that $T = o(n)$. See Heckman and Sedlacek (1985, p. 1088).

estimating $\widehat{\omega}$, is the term that would arise if ω were not estimated. The term reflects typical findings in long panel analysis (i.e., large n, large $\mathbb{T}$), where the standard errors are inversely proportional to the square root of the number $n \times \mathbb{T}$ of observations. The fact that the estimation error of $\widehat{\gamma}$ is dominated by the $O_p\left(\mathbb{T}^{-1/2}\right)$ term indicates that the number of observations is effectively equal to $\mathbb{T}$, i.e., the long panel should be treated as a time series problem for all practical purposes.

This conclusion has two interesting implications. First, the sampling noise due to cross-section variation should be ignored and the "standard" asymptotic variance formulae should generally be avoided in panel data analysis when aggregate shocks are present. We note that Lee and Wolpin's (2006, 2010) standard errors use the standard formula that ignores the $O_p\left(\mathbb{T}^{-1/2}\right)$ term. Second, since in most cases the time-series dimension $\mathbb{T}$ of a panel data set is relatively small, despite the theoretical assumption that it grows to infinity, estimators based on panel data will generally be more imprecise than may be expected from the "large" number $n \times \mathbb{T}$ of observations.[34]

D Asymptotic Distribution and Standard Error Formulas for Examples

In this section, we discuss how the discussion in Section 6 applies to the general equilibrium model. We also present characterizations of the asymptotic distributions for the examples in Sections 3 and 5.

D.1 Standard Error Formula Applied to the General Equilibrium Model

Recall our assumption that the (repeated) cross-sectional data include n i.i.d. observations $\left(w_{i,t}, c^*_{i,t}, l^*_{i,t}, F_{i,t}\right)$ for *working* individuals from two periods $t = 1, 2$. Here, $F_{i,t}$ denotes a dummy

[34]This raises an interesting point. Suppose there is an aggregate time series data set available with which consistent estimation of γ is feasible at the standard rate of convergence. Also suppose that the number of time series observations, say τ, is a lot larger than T. In that case we conjecture that the panel data analysis is strictly dominated by the time series analysis from an efficiency point of view.

variable that is equal to one if the agent chooses $S = F$ in the *previous* period. Recall that we use

$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1}\log w_{i,1}^F - \frac{1}{\bar{n}_2}\sum_{i=1}^{\bar{n}_2}\log w_{i,2}^F = \frac{1}{\widehat{\alpha}_1}\left(\log n_1^F - \log n_2^F\right)$$
$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1}\frac{c_{i,1}^*}{l_{i,1}^*} = w_1^F\frac{\widehat{\sigma}}{1-\widehat{\sigma}}$$

as well as

$$\widehat{\log\nu_s} = \widehat{\alpha}_1\left(\log w_s^F - \log w_s^R\right) - \left(\log n_s^F - \log n_s^R\right). \tag{54}$$

The parameters ϱ and ω^2 can then be consistently estimated by the time-series regression of the following equation:

$$\widehat{\log\nu_{s+1}} = \varrho\widehat{\log\nu_s} + \eta_s. \tag{55}$$

In addition to these equations, we will use the cross section variances of $\log w_{i,1}^F$ and $\log w_{i,1}^R$ to estimate σ_F^2 and σ_R^2. We also have the log likelihood from a sample of n individuals (cross section) is

$$\sum_{i=1}^{n}\left\{F_{i,2}\log\left[1-\Phi\left(\log\left(1-\Theta\right)-\mu\right)\right] + \left(1-F_{i,2}\right)\log\left[\Phi\left(\log\left(1-\Theta\right)-\mu\right)\right]\right\},$$

where Θ is constant across i and given by

$$\Theta \equiv \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2-\sigma_F^2}{2} + \varrho\log\nu_1}{\frac{\sigma\left(\sigma_R^2-\sigma_F^2+\omega^2\right)}{2\alpha_1}}. \tag{56}$$

The moments employed in the estimation of α_1 and σ take the following form:

$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1}\log w_{i,1}^F - \frac{1}{\bar{n}_2}\sum_{i=1}^{\bar{n}_2}\log w_{i,2}^F = \frac{1}{\widehat{\alpha}_1}\left(\log n_1^F - \log n_2^F\right)$$
$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1}\frac{c_{i,1}^*}{l_{i,1}^*} = w_1^F\frac{\widehat{\sigma}}{1-\widehat{\sigma}}.$$

To simplify notation we introduce two redundant parameters δ_1 and δ_2

$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1}\log w_{i,1}^F = \widehat{\delta}_1, \quad \frac{1}{\bar{n}_2}\sum_{i=1}^{\bar{n}_2}\log w_{i,2}^F = \widehat{\delta}_2$$

and understand

$$\widehat{\alpha}_1 = \frac{\log n_1^F - \log n_2^F}{\widehat{\delta}_1 - \widehat{\delta}_2}. \tag{57}$$

Given that our asymptotics are based on $n \to \infty$, we need to express moments in terms of n:

$$\sum_{i=1}^{n} F_{i,1}\left(\log w_{i,1}^F - \delta_1\right) = 0,$$
$$\sum_{i=1}^{n} F_{i,2}\left(\log w_{i,2}^F - \delta_2\right) = 0,$$
$$\sum_{i=1}^{n} F_{i,1}\left(\frac{c_{i,1}^*}{l_{i,1}^*} - w_1^F \frac{\sigma}{1-\sigma}\right) = 0.$$

For the estimation of $\sigma_F^2 = \sigma_\varepsilon^2$, we use the fact that the second moment is the sum of the variance and the square of the first moment and let

$$\sum_{i=1}^{n} F_{i,1}\left(\left(\log w_{i,1}^F\right)^2 - \left(\sigma_F^2 + \delta_1^2\right)\right) = 0.$$

Likewise, for the estimation of σ_R^2,

$$\sum_{i=1}^{n} (1 - F_{i,1})\left(\log w_{i,1}^R - \delta_3\right) = 0,$$
$$\sum_{i=1}^{n} (1 - F_{i,1})\left(\left(\log w_{i,1}^R\right)^2 - \left(\sigma_R^2 + \delta_3^2\right)\right) = 0.$$

For the estimation of the parameters ρ and ω^2, the OLS estimator of ϱ and the corresponding estimator for ω^2 solve:

$$\frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \widehat{\log \nu_s}\left(\widehat{\log \nu_{s+1}} - \widehat{\varrho}\widehat{\log \nu_s}\right) = 0$$

and

$$\frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \left(\widehat{\log \nu_{s+1}} - \widehat{\varrho}\widehat{\log \nu_s}\right)^2 = \widehat{\omega}^2.$$

Replacing for $\widehat{\log \nu_{s+1}}$ and $\widehat{\log \nu_s}$ using equation(54), as well as (57), we obtain the following two moment conditions:

$$\sum_{s=\tau_0+1}^{\tau_0+\tau} \begin{pmatrix} \frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_s^F - \log w_s^R\right) \\ -\left(\log n_t^F - \log n_t^R\right) \end{pmatrix} \times$$
$$\left(\begin{pmatrix} \frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_{s+1}^F - \log w_{s+1}^R\right) \\ -\left(\log n_{s+1}^F - \log n_{s+1}^R\right) \end{pmatrix} - \varrho \begin{pmatrix} \frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_s^F - \log w_s^R\right) \\ -\left(\log n_s^F - \log n_s^R\right) \end{pmatrix}\right) = 0,$$

$$\sum_{s=\tau_0+1}^{\tau_0+\tau} \left(\left(\begin{pmatrix} \frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_{s+1}^F - \log w_{s+1}^R\right) \\ -\left(\log n_{s+1}^F - \log n_{s+1}^R\right) \end{pmatrix} - \varrho \begin{pmatrix} \frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_s^F - \log w_s^R\right) \\ -\left(\log n_s^F - \log n_s^R\right) \end{pmatrix}\right)^2 - \omega^2\right) = 0.$$

For the rest of the parameters, we note that $F_{i,2}$ is chosen with probability $1 - \Phi\left(\log\left(1-\Theta\right) - \mu\right)$ for

$$\Theta = \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_1}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}},$$

so μ can be estimated by Probit MLE, where the FOC can be shown to be

$$0 = \sum_{i=1}^{n} \left\{F_{i,2} - \left[1 - \Phi\left(\log\left(1-\Theta\right) - \mu\right)\right]\right\},$$

where

$$\begin{aligned}\Theta &= \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_1}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}} \\ &= \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho\left(\frac{\log n_1^F - \log n_2^F}{\delta_1-\delta_2}\left(\log w_1^F - \log w_1^R\right) - \left(\log n_1^F - \log n_1^R\right)\right)}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2}\frac{\delta_1-\delta_2}{\log n_1^F - \log n_2^F}}.\end{aligned}$$

Here, we used the fact that

$$\begin{aligned}\log \nu_1 &= \alpha_1\left(\log w_1^F - \log w_1^R\right) - \left(\log n_1^F - \log n_1^R\right) \\ \alpha_1 &= \frac{\log n_1^F - \log n_2^F}{\delta_1 - \delta_2}.\end{aligned}$$

Based on the previous discussion, we can now present moments in the form of (33) and (34). In our case, $\log \nu_1$ is estimated with the aid of aggregate variables, so we have $\beta = \theta = (\mu, \delta_1, \delta_2, \sigma, \delta_3, \sigma_F^2, \sigma_R^2)'$ and $\rho = (\varrho, \omega^2)'$. We see that the cross sectional moments are

$$\frac{1}{n}\sum_{i=1}^{n} F_{i,1}\left(\log w_{i,1}^F - \delta_1\right) = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n} F_{i,2}\left(\log w_{i,2}^F - \delta_2\right) = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n} F_{i,1}\left(\frac{c_{i,1}^*}{l_{i,1}^*} - w_1^F \frac{\sigma}{1-\sigma}\right) = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n} \left\{F_{i,2} - \left[1 - \Phi\left(\log\left(1-\Theta\right) - \mu\right)\right]\right\} = 0,$$

and

$$\sum_{i=1}^{n} F_{i,1}\left(\left(\log w_{i,1}^F\right)^2 - \left(\sigma_F^2 + \delta_1^2\right)\right) = 0,$$
$$\sum_{i=1}^{n} \left(1 - F_{i,1}\right)\left(\log w_{i,1}^R - \delta_3\right) = 0,$$
$$\sum_{i=1}^{n} \left(1 - F_{i,1}\right)\left(\left(\log w_{i,1}^R\right)^2 - \left(\sigma_R^2 + \delta_3^2\right)\right) = 0,$$

where

$$\Theta = \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho\left(\frac{\log n_1^F - \log n_2^F}{\delta_1 - \delta_2}\left(\log w_1^F - \log w_1^R\right) - \left(\log n_1^F - \log n_1^R\right)\right)}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2} \frac{\delta_1 - \delta_2}{\log n_1^F - \log n_2^F}},$$

and the time series moments are

$$\frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \log \nu_s \left(\log \nu_{s+1} - \varrho \log \nu_s\right) = 0,$$
$$\frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \left(\left(\log \nu_{s+1} - \varrho \log \nu_s\right)^2 - \omega^2\right) = 0,$$

where

$$\log \nu_s = \frac{\log n_1^F - \log n_2^F}{\delta_1 - \delta_2} \left(\log w_s^F - \log w_s^R \right) - \left(\log n_s^F - \log n_s^R \right).$$

Letting

$$f_{\theta,i}(\theta, \rho) = \begin{bmatrix} F_{i,1} \left(\log w_{i,1}^F - \delta_1 \right) \\ F_{i,1} \left(\frac{c_{i,1}^*}{l_{i,1}^*} - w_1^F \frac{\sigma}{1-\sigma} \right) \\ F_{i,1} \left(\left(\log w_{i,1}^F \right)^2 - (\sigma_F^2 + \delta_1^2) \right) \\ (1 - F_{i,1}) \left(\log w_{i,1}^R - \delta_3 \right) \\ (1 - F_{i,1}) \left(\left(\log w_{i,1}^R \right)^2 - (\sigma_R^2 + \delta_3^2) \right) \\ F_{i,2} \left(\log w_{i,2}^F - \delta_2 \right) \\ F_{i,2} - [1 - \Phi(\log(1 - \Theta) - \mu)] \end{bmatrix}, \tag{58}$$

and

$$g_{\rho,s}(\beta, \rho) = \begin{bmatrix} \log \nu_s (\log \nu_{s+1} - \varrho \log \nu_s) \\ (\log \nu_{s+1} - \varrho \log \nu_s)^2 - \omega^2 \end{bmatrix}, \tag{59}$$

we can compute

$$\hat{\Omega}_f = \frac{1}{n} \sum_{i=1}^{n} f_{\theta,i} f'_{\theta,i}$$

and

$$\hat{\Omega}_g = \tau^{-1} \sum_{s=\tau_0+1}^{\tau_0+\tau} g_{\rho,s} g'_{\rho,s}$$

and

$$\hat{W} = \begin{bmatrix} \frac{1}{n}\hat{\Omega}_f & 0 \\ 0 & \frac{1}{\tau}\hat{\Omega}_g \end{bmatrix}. \tag{60}$$

We are now ready to describe the five steps required in the computation of test statistics and confidence intervals for the general equilibrium model. As a first step, let $\theta = \beta = (\mu, \delta_1, \delta_2, \sigma, \delta_3, \sigma_F^2, \sigma_R^2)'$ and $\rho = (\varrho, \omega^2)'$. Observe that the aggregate shock is not in the set of estimated parameters, since the general equilibrium model implies that $\log \nu_s = \alpha_1 \left(\log w_s^F - \log w_s^R \right) - \left(\log n_s^F - \log n_s^R \right)$. In the second, third, and fourth steps compute the matrices $\mathbf{A}$, $\hat{\Omega}_f$, $\hat{\Omega}_g$, and W using the vectors of moments $f_{\theta,i}$ and $g_{\rho,s}$ derived above. In the last step, calculate the variance matrix $V = \mathbf{A}^{-1} W (\mathbf{A}')^{-1}$ and form related t-ratios and confidence intervals.

D.2 Limiting Distributions

We first consider the portfolio choice problem in Section 3. In this example, the time series log likelihood is given by

$$\tau^{-1}\sum_{s=\tau_0+1}^{\tau_0+\tau}\log\left(\phi\left(\left(\nu_s-\mu\right)/\sigma_\nu\right)/\sigma_\nu\right),$$

where ϕ is the PDF of $N(0,1)$. The likelihood is maximized that $\hat{\mu}=\tau^{-1}\sum_{s=\tau_0+1}^{\tau_0+\tau}\nu_s$ and $\hat{\sigma}_\nu^2=\tau^{-1}\sum_{s=\tau_0+1}^{\tau_0+\tau}\left(\nu_s-\hat{\mu}\right)^2$. The cross-sectional likelihood is given by

$$n^{-1}\sum_{i=1}^{n}\log\left(\phi\left(\left(u_{i1}-\nu_1\right)/\sigma_\epsilon\right)/\sigma_\epsilon\right)+n^{-1}\sum_{i=1}^{n}\log\left(\phi\left(\left(\alpha_{i1}-\alpha\right)/\sigma_e\right)/\sigma_e\right),$$

where $\alpha=\left(\delta\left(\sigma_\epsilon^2+\sigma_\nu^2\right)+r-\mu\right)/\delta\left(\sigma_\epsilon^2+\sigma_\nu^2\right)$. For given values of $\mu, r,$and σ_ν^2 there is a one-to-one mapping between the parameters $\left(\delta,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$ and $\left(\alpha,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$. Maximizing the likelihood with respect to $\left(\delta,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$ is thus equivalent to maximizing the likelihood with respect to $\left(\alpha,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$ and then solving for $\left(\delta,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$. The maximizer for $\left(\alpha,\sigma_\epsilon^2,\sigma_e^2,\nu_1\right)$ is the standard MLE of the normal distribution for mean and variance, $\hat{\nu}_1=n^{-1}\sum_{i=1}^{n}u_{i1}$, $\hat{\alpha}=n^{-1}\sum_{i=1}^{n}\alpha_{i1}$, $\hat{\sigma}_\epsilon^2=n^{-1}\sum_{i=1}^{n}\left(u_{i1}-\hat{\nu}_1\right)^2$ and $\hat{\sigma}_e=n^{-1}\sum_{i=1}^{n}\left(\alpha_{i1}-\hat{\alpha}\right)^2$. The limiting distributions of these estimators are given by

$$\tau^{1/2}\begin{pmatrix}\hat{\mu}-\mu\\ \hat{\sigma}_\nu^2-\sigma_\nu^2\end{pmatrix}\rightarrow_d N\left(0,\begin{bmatrix}\sigma_\nu^2 & 0\\ 0 & 2\sigma_\nu^2\end{bmatrix}\right),$$

and

$$n^{1/2}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\sigma}_\epsilon^2-\sigma_\epsilon^2\\ \hat{\sigma}_e^2-\sigma_e^2\\ \hat{\nu}_1-\nu_1\end{pmatrix}\rightarrow_d N\left(0,\begin{bmatrix}\sigma_e^2 & 0 & 0 & 0\\ 0 & 2\sigma_\epsilon^2 & 0 & 0\\ 0 & 0 & 2\sigma_e^2 & 0\\ 0 & 0 & 0 & \sigma_\epsilon^2\end{bmatrix}\right).$$

From the results in Hahn, Kuersteiner, and Mazzocco (2016) the convergence of the two vectors is joint, with asymptotic independence between cross-section and time series parameters, and stable with respect to ν_1. However, because of the particularly simple nature of the model the limiting distributions are conventional Gaussian limits with fixed variances. To obtain the limiting distribution of $\hat{\delta}$ one now simply applies the delta method and the continuous mapping theorem.

More specifically, we have $\hat{\delta} = (\hat{\mu} - r) / ((\hat{\sigma}_\epsilon^2 + \hat{\sigma}_\nu^2)(1-\hat{\alpha}))$ and

$$n^{-1/2}\left(\hat{\delta} - \delta\right) = \frac{\mu - r}{(\sigma_\epsilon^2 + \sigma_\nu^2)(1-\alpha)^2} n^{1/2}(\hat{\alpha} - \alpha) - \frac{\mu - r}{(\sigma_\epsilon^2 + \sigma_\nu^2)^2 (1-\alpha)} n^{1/2}\left(\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2\right) \tag{61}$$
$$+ \frac{1}{(\sigma_\epsilon^2 + \sigma_\nu^2)(1-\alpha)} \sqrt{\frac{n}{\tau}} \tau^{1/2} (\hat{\mu} - \mu) - \frac{\mu - r}{(\sigma_\epsilon^2 + \sigma_\nu^2)^2 (1-\alpha)} \sqrt{\frac{\tau}{n}} \tau^{1/2} \left(\hat{\sigma}_\nu^2 - \sigma_\nu^2\right) + o_p(1),$$

leading to a limiting distribution of $\hat{\delta}$ given by

$$n^{-1/2}\left(\hat{\delta} - \delta\right) \to_d N\left(0, \frac{2(1-\alpha)^2 (\mu - r)^2 (\sigma_\epsilon^2 + \kappa\sigma_\nu^2) + (\sigma_\epsilon^2 + \sigma_\nu^2)^2 \left((\mu - r)^2 \sigma_e^2 + (1-\alpha)^2 \kappa \sigma_\nu^2\right)}{(1-\alpha)^4 (\sigma_\epsilon^2 + \sigma_\nu^2)^4}\right),$$

where $\kappa = \lim \frac{n}{\tau}$ and the variance formula uses the fact that the four components in (61) are asymptotically independent. The formula for the variance is indicative of the fact that first step estimation of the time series parameters can be ignored if τ is much larger than n, such that κ is close to zero. However, this is an unlikely scenario given that cross-sectional samples tend to be quite large.

We now consider the general equilibrium example. It is useful to analyze the form of the limiting distribution of a set of GMM estimators based on f and g. Define the empirical moment functions as

$$h_n(\theta, \rho) = n^{-1} \sum_{i=1}^{n} f_{\theta,i}(\theta, \rho), \quad k_\tau(\beta, \rho) = \tau^{-1} \sum_{t=\tau_0+1}^{\tau_0+\tau} g_{\rho,t}(\beta, \rho),$$

and the moment based criterion functions $F_n(\theta, \rho) = -h_n(\theta, \rho)' \hat{\Omega}_y^{-1} h_n(\theta, \rho)$ and $G_\tau(\beta, \rho) = -k_\tau(\beta, \rho)' \hat{\Omega}_\nu^{-1} k_\tau(\beta, \rho)$. The estimators then are defined as the solution $\left(\hat{\theta}, \hat{\rho}\right)$ to

$$\frac{\partial F_n\left(\hat{\theta}, \hat{\rho}\right)}{\partial \theta} = 0$$
$$\frac{\partial G_\tau\left(\hat{\beta}, \hat{\rho}\right)}{\partial \rho} = 0.$$

Because the GMM estimators are exactly identified in our example these equations reduce to

$$h_n\left(\hat{\theta},\hat{\rho}\right)=0$$
$$k_\tau\left(\hat{\beta},\hat{\rho}\right)=0.$$

We focus on the just identified case and refer the reader to our companion paper Hahn, Kuersteiner and Mazzocco (2016) for a general treatment. The limiting distribution of $\hat{\theta},\hat{\rho}$ depends on the joint limiting distribution of $h_n(\theta_0,\rho_0)$ and $k_\tau(\beta_0,\rho_0)$.

Recall $\log w_{it}^F=\alpha_1^{-1}\left(\log n_t^F+\log\sigma+\log T-\alpha_0\right)+\varepsilon_{it}^F$ such that

$$\delta_1=\alpha_1^{-1}\left(\log n_1^F+\log\sigma+\log T-\alpha_0\right)-\frac{\sigma_F^2}{2}.$$

Similarly, let $\delta_2=\alpha_1^{-1}\left(\log n_2^F+\log\sigma+\log T-\alpha_0\right)-\sigma_F^2/2$,

$$\delta_3=\alpha_1^{-1}\left(\log n_1^R+\log\sigma+\log T-\alpha_0-\log\nu_1\right)-\frac{\sigma_R^2}{2}$$

and define $p(\Theta)=\Phi(\log(1-\Theta)-\mu)$. Let $\mathcal{C}$ be the σ-field generated by $\log n_1^R,\log n_1^F,\log n_2^F$ and $\log\nu_1$ such that Θ, w_1^F, δ_1,δ_2 and δ_3 are measurable with respect to $\mathcal{C}$. A formal definition of $\mathcal{C}$-stable convergence is due to Renyi (1963).

Definition 1 *Let Z^n and Z be random variables defined on a joint probability space $(\Omega,\mathcal{F},P)$ taking values in $\mathbb{R}^d$ and let $\mathcal{C}$ be as sub-sigma field of $\mathcal{F}$. The sequence Z^n converges $\mathcal{C}$-stably to Z if for all bounded ζ measurable with respect to $\mathcal{C}$ it follows that*

$$E\left[\zeta\exp\left(itZ^n\right)\right]\to E\left[\zeta\exp\left(itZ\right)\right]$$

for $i=\sqrt{-1}$ and $t\in\mathbb{R}^d$.

The exact form of Definition 1 is due to Aldous and Eagleson (1978) who show that it is equivalent to the joint weak convergence of Z^n and ζ.

Based on the theory in our companion paper, the moment functions converge jointly and stably to independent mixed Gaussian limits

$$n^{1/2}h_n(\theta_0,\rho_0)\to_d \Omega_f^{1/2}\xi_h\sim N(0,\Omega_f)\quad (\mathcal{C}\text{-stably}),$$

where $\xi_h\sim N(0,I)$ and is independent of any $\mathcal{C}$-measurable random variable,

$$\Omega_{f,1}=\begin{bmatrix} p\left(\bar{\Theta}_1\right)\sigma_F^2 & p\left(\bar{\Theta}_1\right)\frac{w_1^F\sigma}{1-\sigma}\sigma_F^2 & 2\delta_1\sigma_F^2 \\ p\left(\bar{\Theta}_1\right)\frac{w_1^F\sigma}{1-\sigma}\sigma_F^2 & p\left(\bar{\Theta}_1\right)\left(\frac{w_1^F\sigma}{1-\sigma}\right)^2\left(e^{\sigma_F^2}-1\right) & p\left(\bar{\Theta}_1\right)\frac{w_1^F\sigma}{1-\sigma}(2\delta_1+1)\sigma_\epsilon^2 \\ 2\delta_1\sigma_F^2 & p\left(\bar{\Theta}_1\right)\frac{w_1^F\sigma}{1-\sigma}(2\delta_1+1)\sigma_F^2 & p\left(\bar{\Theta}_1\right)(2\sigma_\epsilon^4+4\delta_1^2\sigma_F^2)\end{bmatrix},$$

$$\Omega_{f,2}=\begin{bmatrix}\left(1-p\left(\bar{\Theta}_1\right)\right)\sigma_R^2 & 2\delta_3\sigma_R^2 \\ 2\delta_3\sigma_R^2 & \left(1-p\left(\bar{\Theta}_1\right)\right)(2\sigma_R^4+4\delta_3^2\sigma_R^2)\end{bmatrix},$$

$$\Omega_{f,3}=\begin{bmatrix} p\left(\bar{\Theta}_1\right)\sigma_\epsilon^2 & 0 \\ 0 & p\left(\bar{\Theta}_2\right)\left(1-p\left(\bar{\Theta}_2\right)\right)\end{bmatrix}$$

and

$$\Omega_f=\begin{bmatrix}\Omega_{f,1} & 0 & 0\\ 0 & \Omega_{f,2} & 0 \\ 0 & 0 & \Omega_{f,3}\end{bmatrix}.$$

Here, we let

$$\bar{\Theta}_t\equiv\frac{\log\left(\frac{n_t^F}{n_t^R}\right)+\frac{\left(\pi_2^2-1\right)\sigma_\varepsilon^2}{2}+\varrho\log\nu_{t-1}}{\frac{\sigma\left(\sigma_R^2-\sigma_F^2+\omega^2\right)}{2\alpha_1}}$$

for clarity. For the time series sample it is straight forward to see that under suitable regularity conditions

$$\tau^{1/2}k_\tau(\beta_0,\rho_0)\to_d\Omega_g^{1/2}\xi_k\sim N(0,\Omega_g)\quad(\mathcal{C}\text{-stably}),$$

where $\xi_k\sim N(0,I)$ and independent of any $\mathcal{C}$-measurable random variable and

$$\Omega_g=\begin{bmatrix}\frac{\omega^4}{1-\varrho_0^2} & 0 \\ 0 & 2\omega^4\end{bmatrix}.$$

The results in Hahn, Kuersteiner and Mazzocco (2016) imply that ξ_h and ξ_k are independent

Gaussian random variables conditional on $\mathcal{C}$. The explicit formulas make clear that in this model the limiting variance does depend on macro variables including common shocks and other observables. Since these variables remain random in the limit as n and τ tend to infinity, the resulting limiting distribution is mixed Gaussian and the convergence to the limit is joint with the macro variables or $\mathcal{C}$-stable. The later is important because the influence matrix A, as we show below, also depends on these same macro variables.

Next compute the limits

$$A_{f,\theta} = \operatorname{plim} n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\theta_0, \rho_0\right)}{\partial \theta'}, \qquad A_{f,\rho} = \operatorname{plim} n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\theta_0, \rho_0\right)}{\partial \rho'},$$

$$A_{g,\theta} = \operatorname{plim} \tau^{-1} \sum_{t=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,t}\left(\beta_0, \rho_0\right)}{\partial \theta'}, \qquad A_{g,\rho} = \operatorname{plim} \tau^{-1} \sum_{t=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,t}\left(\beta_0, \rho_0\right)}{\partial \rho'}.$$

First, letting $\dot{p}\left(\Theta\right) = \phi\left(\log\left(1-\Theta\right) - \mu\right)$ where ϕ is the PDF of $N\left(0,1\right)$,

$$A_{f,\theta} = \begin{bmatrix} 0 & -p\left(\bar{\Theta}_1\right) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{-w_1^F}{(1-\sigma)^2} p\left(\bar{\Theta}_1\right) & 0 & 0 & 0 \\ 0 & -2\delta_1 p\left(\bar{\Theta}_1\right) & 0 & 0 & 0 & -p\left(\bar{\Theta}_1\right) & 0 \\ 0 & 0 & 0 & 0 & -\left(1-p\left(\bar{\Theta}_1\right)\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & -2\delta_3\left(1-p\left(\bar{\Theta}_1\right)\right) & 0 & -\left(1-p\left(\bar{\Theta}_1\right)\right) \\ 0 & 0 & -p\left(\bar{\Theta}_2\right) & 0 & 0 & 0 & 0 \\ -\dot{p}\left(\bar{\Theta}_2\right) & -\frac{\dot{p}\left(\bar{\Theta}_2\right)}{1-\bar{\Theta}_2}\frac{\partial\bar{\Theta}_2}{\partial\delta_1} & -\frac{\dot{p}\left(\bar{\Theta}_2\right)}{1-\bar{\Theta}_2}\frac{\partial\bar{\Theta}_2}{\partial\delta_2} & -\frac{\dot{p}\left(\bar{\Theta}_2\right)}{1-\bar{\Theta}_2}\frac{\partial\bar{\Theta}_2}{\partial\sigma} & 0 & -\frac{\dot{p}\left(\bar{\Theta}_2\right)}{1-\bar{\Theta}_2}\frac{\partial\bar{\Theta}_2}{\partial\sigma_F^2} & -\frac{\dot{p}\left(\bar{\Theta}_2\right)}{1-\bar{\Theta}_2}\frac{\partial\bar{\Theta}_2}{\partial\sigma_R^2} \end{bmatrix},$$

Next, consider the two cross-derivative terms where the first one is given by

$$A_{f,\rho} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\frac{\dot{p}\left(\bar{\Theta}\right)}{1-\bar{\Theta}}\frac{\partial\bar{\Theta}}{\partial\varrho} & -\frac{\dot{p}\left(\bar{\Theta}\right)}{1-\bar{\Theta}}\frac{\partial\bar{\Theta}}{\partial\omega^2} \end{bmatrix}.$$

Next note that

$$\log \nu_s = \frac{\log n_1^F - \log n_2^F}{\delta_1 - \delta_2} \left(\log w_s^F - \log w_s^R\right) - \left(\log n_s^F - \log n_s^R\right)$$

such that $\partial \log \nu_s / \partial \theta$ is non-zero for elements δ_1 and δ_2. For $\log \nu_s \left(\log \nu_{s+1} - \varrho \log \nu_s\right)$ the derivative $\left(\partial \log \nu_s / \partial \theta\right) \left(\log \nu_{s+1} - \varrho_0 \log \nu_s\right)$ has zero expectation because $\left(\log \nu_{s+1} - \varrho_0 \log \nu_s\right) = \eta_s$. For $\left(\log \nu_{s+1} - \varrho \log \nu_s\right)^2 - \omega^2$ we obtain partial derivatives equal to $2\eta_s \left(\partial \log \nu_{s+1} / \partial \theta - \varrho \partial \log \nu_s / \partial \theta\right)$. Since η_s is orthogonal to all data in $\log \nu_s$ it follows that $E\left[\eta_s \left(\partial \log \nu_{s+1} / \partial \theta - \varrho \partial \log \nu_s / \partial \theta\right)\right] = E\left[\eta_s \partial \log \nu_{s+1} / \partial \theta\right]$. Under suitable regularity conditions it then follows that sample averages converge to these expectations, leading to

$$A_{g,\theta} = \begin{bmatrix} 0 & E\left[\log \nu_s \left(\frac{\partial \log \nu_{s+1}}{\partial \delta_1} - \varrho \frac{\partial \log \nu_s}{\partial \delta_1}\right)\right] & E\left[\log \nu_s \left(\frac{\partial \log \nu_{s+1}}{\partial \delta_2} - \varrho \frac{\partial \log \nu_s}{\partial \delta_2}\right)\right] & 0 & 0 & 0 & 0 \\ 0 & 2E\left[\eta_s \frac{\partial \log \nu_{s+1}}{\partial \delta_1}\right] & 2E\left[\eta_s \frac{\partial \log \nu_{s+1}}{\partial \delta_2}\right] & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Finally, straight forward calculations show that under suitable regularity conditions ensuring a law of large numbers for an autoregressive process the limits in $A_{g,\rho}$ are given by

$$A_{g,\rho} = \begin{bmatrix} -\frac{\omega^2}{1-\varrho^2} & 0 \\ 0 & -1 \end{bmatrix}.$$

The limiting distribution of $\hat{\theta}$ is a consequence of Hahn, Kuersteiner and Mazzocco (2016), Theorem 2 and Corollary 2. Using the notation developed here we have

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} -A^{f,\theta} \Omega_f^{1/2} \xi_h - \sqrt{\kappa} A^{g,\rho} \Omega_g^{1/2} \xi_k \text{ } (\mathcal{C}\text{-stably}),$$

where

$$A^{f,\theta} = A_{f,\theta}^{-1} + A_{f,\theta}^{-1} A_{f,\rho} \left(A_{g,\rho} - A_{g,\theta} A_{f,\theta}^{-1} A_{f,\rho}\right)^{-1} A_{g,\theta} A_{f,\theta}^{-1}$$
$$A^{g,\rho} = -A_{f,\theta}^{-1} A_{f,\rho} \left(A_{g,\rho} - A_{g,\theta} A_{f,\theta}^{-1} A_{f,\rho}\right)^{-1}.$$

The limiting distribution of $\hat{\theta}$ is mixed Gaussian $N\left(0, \Omega_\theta\right)$, with random weight matrix $\Omega_\theta = A^{f,\theta} \Omega_f A^{f,\theta\prime} + \kappa A^{g,\rho} \Omega_g A^{g,\rho\prime}$ where we have shown how the elements of A and Ω_f depend on macro variables and unobserved macro shocks. Similarly, the limiting distribution of $\hat{\rho}$ is also mixed Gaussian and can be derived in a similar fashion.

E Proof of (31)

Suppose that our econometrician tries to estimate μ using only cross-section data sets misspecifies the model and assumes that the difference in the labor demand functions of the two types of firms is not due to the aggregate shock, but to different intercepts, i.e.,

$$\log H_{t+1}^{D,F} = \alpha_0 + \alpha_1 \log w_{t+1}^F$$
$$\log H_{t+1}^{D,R} = \alpha_0' + \alpha_1 \log w_{t+1}^R$$

with $\alpha_0 \neq \alpha_0'$. The equilibrium wages are then

$$\begin{aligned} \log w_{t+1}^F &= \frac{\log n_{t+1}^F + \log \sigma + \log T - \alpha_0}{\alpha_1}, \\ \log w_{t+1}^R &= \frac{\log n_{t+1}^R + \log \sigma + \log T - \alpha_0'}{\alpha_1}, \end{aligned} \tag{62}$$

and as a consequence, equation (43) is changed to

$$\begin{aligned} &\left(n_{t+1}^F\right)^{\sigma(1-\gamma)/\alpha_1} \left[\left(\left(\frac{1}{e^{\alpha_0}}\right)^{1/\alpha_1}\right)^{\sigma}\right]^{1-\gamma} E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}^F\right)\right] \\ &\geq \left(n_{t+1}^R\right)^{\sigma(1-\gamma)/\alpha_1} \left[\left(\left(\frac{1}{e^{\alpha_0'}}\right)^{1/\alpha_1}\right)^{\sigma}\right]^{1-\gamma} E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}^R\right)\right]. \end{aligned}$$

Note that

$$\begin{aligned} E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}^F\right)\right] &= \exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)}{2}\sigma_\varepsilon^2\right)\exp\left(\frac{\left(\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\right)^2}{2}\sigma_F^2\right), \\ E\left[\exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}^R\right)\right] &= \exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)}{2}\pi_2^2\sigma_\varepsilon^2\right)\exp\left(\frac{\left(\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\right)^2}{2}\sigma_R^2\right), \end{aligned}$$

and

$$\left[\left(\left(\frac{1}{e^{\alpha_0}}\right)^{1/\alpha_1}\right)^{\sigma}\right]^{1-\gamma}\exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)}{2}\sigma_F^2\right)=\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0\right),$$

$$\left[\left(\left(\frac{1}{e^{\alpha_0'}}\right)^{1/\alpha_1}\right)^{\sigma}\right]^{1-\gamma}\exp\left(-\frac{\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)}{2}\sigma_R^2\right)=\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0'\right),$$

where

$$\widetilde{\alpha}_0=\alpha_0+\frac{1}{2}\sigma_F^2=\alpha_0-E\left[\varepsilon_{t+1}^F\right],\quad \widetilde{\alpha}_0'=\alpha_0'+\frac{1}{2}\sigma_R^2=\alpha_0'-E\left[\varepsilon_{t+1}^R\right].$$

Therefore, the econometrician will conclude that F is chosen if

$$\begin{aligned}&\left(n_{t+1}^F\right)^{\sigma(1-\gamma)/\alpha_1}\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0\right)\\&\qquad\geq\left(n_{t+1}^R\right)^{\sigma(1-\gamma)/\alpha_1}\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0'\right)\exp\left(\frac{\left(\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\right)^2}{2}\left(\sigma_R^2-\sigma_F^2\right)\right)\end{aligned}$$

when $1-\gamma>0$, and

$$\begin{aligned}&\left(n_{t+1}^F\right)^{\sigma(1-\gamma)/\alpha_1}\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0\right)\\&\qquad\leq\left(n_{t+1}^R\right)^{\sigma(1-\gamma)/\alpha_1}\exp\left(-\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\widetilde{\alpha}_0'\right)\exp\left(\frac{\left(\sigma\left(1-\gamma\right)\left(1/\alpha_1\right)\right)^2}{2}\left(\sigma_R^2-\sigma_F^2\right)\right)\end{aligned}$$

when $1-\gamma<0$. This implies that F is chosen if

$$\gamma\geq 1-\frac{\log\left(\frac{n_{t+1}^F}{n_{t+1}^R}\right)+\left(\widetilde{\alpha}_0'-\widetilde{\alpha}_0\right)}{\frac{\sigma\left(\pi_2^2-1\right)\sigma_\varepsilon^2}{2\alpha_1}}. \tag{63}$$

Note that

$$\widetilde{\alpha}_0'-\widetilde{\alpha}_0=\alpha_0'-\alpha_0+\frac{1}{2}\left(\sigma_R^2-\sigma_F^2\right).$$

We now argue that $\alpha_0'-\alpha_0$ above should be understood to be equal to $\log v_{t+1}$. Note that the econometrician can estimate α_1 consistently using equation (26), which is based on cross-section variation. The econometrician can also estimate $\alpha_0'-\alpha_0$ consistently by $\widehat{\alpha}_1\left(\log w_{t+1}^F-\log w_{t+1}^R\right)-\left(\log n_{t+1}^F-\log n_{t+1}^R\right)$. Comparing with (28), we conclude that the econometrician's estimator is exactly equal to our earlier estimator of $\log \nu_{t+1}$. This is a natural consequence of the nature of the econometrician's misspecification, who assumes that the difference in the equilibrium wages in (62) reflects the difference of intercepts of the labor demand functions. However, this assumption is incorrect and the difference of the intercepts is due to the aggregate shock, i.e, $\alpha_0' = \alpha_0 + \log \upsilon_{t+1}$.

It follows that the econometrician's conclusion (63) above can be equivalently written with $\alpha_0' - \alpha_0$ replaced by $\log \upsilon_{t+1}$, which establishes (31).

F Censored versus Truncated Results

As mentioned in the main text, to perform the Monte Carlo exercise we have to deal with a technical issue. The estimation of the risk aversion parameter μ in the general equilibrium model requires the computation of $\log(1-\Theta)$ where

$$\Theta \equiv \frac{\log\left(\frac{n_2^F}{n_2^R}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_1}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}}.$$

In the model, Θ is always smaller than 1 and, hence, $\log(1-\Theta)$ is always well defined. In the estimation of μ, however, the true parameters included in Θ are replaced with their estimated values. In some of the Monte Carlo repetitions, the randomness of the estimated parameters generates values of Θ that are greater than 1, which implies that $\log(1-\Theta)$ is not well define. We deal with this issue by presenting two sets of results. A first set in which we only use Monte Carlo runs in which $\Theta < 1$. We will refer to these results as the "truncated" results. A second set in which we set $\Theta = 0.99$ if $\Theta > 1$ and report our findings using all the Monte Carlo runs. We will refer to the second set as the "censored" set. With the results, we also report the number of simulations in which $\Theta > 1$. An examination of the probability of choosing education F clarifies that the censored set tends to bias the estimates of μ downward: by setting Θ closer to 1, the MLE estimator of μ tends to minus infinity. The truncated set may therefore provide a more accurate description of the true bias. But the censored set is also informative because it documents the potential effect of replacing the true parameters of the model with their estimates in the estimation of parameters that are affected by both cross-sectional and time-series variation.

This issue is even more significant when the risk aversion parameter is estimated using the misspecified model. In that case, Θ can be greater than 1 for two different reasons. First, as in the general equilibrium model, the true parameters are replaced by their estimated counterparts. Second, Θ is misspecified and, hence, there is no reason to expect that it satisfies the theoretical restriction $\Theta < 1$. We therefore expect the downward bias for the misspecified model in the censored results and the number of cases in which $\Theta > 1$ to be larger than in the general equilibrium model.

Tables 6 and 7 compare the results obtained using the censored sample with the results obtained using the truncated sample. There are three patterns worth highlighting. First, when the censored sample is used, as expected, the average of the estimated risk aversion parameter obtained employing our proposed method is always lower. Second, with our proposed method the number of cases in which $\Theta > 1$ decreases with the length of the time-series, since the persistence and the variance of the aggregate shocks are estimated more precisely. This suggests that it is important to employ a long time-series of aggregate data to avoid situations in which the estimated parameters are incompatible with the structure of the model. Lastly, as expected, when we use the misspecified model, the number of cases in which $\Theta > 1$ is much larger and the misspecification bias goes from being positive to being negative.

G Proof of $\Theta_t < 1$.

We first prove $\Theta_t < 1$ under the assumption that an equilibrium exists. We then prove the existence of a unique equilibrium.

If an equilibrium exists for the general equilibrium model developed in Section 5, it is straightforward to prove that $\Theta_t < 1$. Suppose $\Theta_t \geq 1$. Since $\gamma_i > 0$, we have

$$1 - \frac{\log\left(\frac{n^F_{t+1}}{n^R_{t+1}}\right) + \frac{\sigma_R^2 - \sigma_F^2}{2} + \varrho \log \nu_t}{\frac{\sigma\left(\sigma_R^2 - \sigma_F^2 + \omega^2\right)}{2\alpha_1}} = 1 - \Theta_t \leq 0 < \gamma_i.$$

Equation (25) then implies that every person will choose the flexible education. As a consequence, $n^R_{t+1} = 0$ and $\log\left(\frac{n^F_{t+1}}{n^R_{t+1}}\right) = \infty$. Hence, $\Theta_t = -\infty$ because $\alpha_1 < 0$, which contradicts the initial assumption that $\Theta_t \geq 1$.

We now show that an equilibrium exists. Let Λ denote the CDF of γ_i. Note that $\Lambda(t) = 0$ for all $t \leq 0$ because of the assumption that $\gamma_i > 0$. Equation (25) implies that the probability that education F is chosen (the proportion of workers who chose F) is equal to $1 - \Lambda(\max(0, 1 - \Theta_t))$. It follows that the proportion of workers who chose R is equal to $\Lambda(\max(0, 1 - \Theta_t))$. As a consequence, in equilibrium, the ratio n^F_{t+1} / n^R_{t+1} is a fixed point of the equality

$$\frac{n^F_{t+1}}{n^R_{t+1}} = \frac{1 - \Lambda(\max(0, 1 - \Theta_t))}{\Lambda(\max(0, 1 - \Theta_t))},$$

where we note that Θ_t is a function of n^F_{t+1} / n^R_{t+1}. Let $x = n^F_{t+1} / n^R_{t+1}$, and write the equilibrium condition as

$$x = \frac{1 - \Lambda(\max(0, 1 - \Theta_t(x)))}{\Lambda(\max(0, 1 - \Theta_t(x)))}. \tag{64}$$

The left hand side is straightforwardly a monotonically increasing function of x. Consider now the right hand side. As x increases from 0 to ∞, $\log\left(\frac{n^F_{t+1}}{n^R_{t+1}}\right) = \log x$ monotonically increases from $-\infty$ to ∞. As a consequence, since $\alpha_1 < 0$, $1 - \Theta_t(x)$ monotonically increases from $-\infty$ to ∞ as x increases from 0 to ∞. Hence, $\max(0, 1 - \Theta_t(x))$ monotonically increases from 0 to ∞. Because Λ is the CDF of a positive valued random variable γ_i, $\Lambda(\max(0, 1 - \Theta_t(x)))$ monotonically increases from 0 to 1 and, as a consequence, $(1 - \Lambda(\max(0, 1 - \Theta_t(x)))) / \Lambda(\max(0, 1 - \Theta_t(x)))$ monotonically decreases from ∞ to 0, as x increases from 0 to ∞. We can therefore conclude that, as x increases from 0 to ∞, the left hand side of (64) montonically increases from 0 to ∞, while the right hand side monotonically decreases from ∞ to 0. Hence, the model has a unique equilibrium.

Table 6: Monte Carlo Results, Parameter Estimates For Correct Model

	Censored Results		*Truncated Results*	
True Parameter	Estimate	Cov. Prob.	Estimate	N. Cases
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 25				
Log Risk Aversion Mean: $\mu = 0.2$	**0.057**	**0.893**	**0.161**	119/5000
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 50				
Log Risk Aversion Mean: $\mu = 0.2$	**0.116**	**0.914**	**0.172**	67/5000
Cross-sectional Sample Size: 2,500, Time-series Sample Size: 100				
Log Risk Aversion Mean: $\mu = 0.2$	**0.126**	**0.933**	**0.173**	59/5000
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 25				
Log Risk Aversion Mean: $\mu = 0.2$	**0.071**	**0.892**	**0.175**	118/5000
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 50				
Log Risk Aversion Mean: $\mu = 0.2$	**0.132**	**0.912**	**0.183**	62/5000
Cross-sectional Sample Size: 5,000, Time-series Sample Size: 100				
Log Risk Aversion Mean: $\mu = 0.2$	**0.145**	**0.928**	**0.173**	36/5000
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 25				
Log Risk Aversion Mean: $\mu = 0.2$	**0.078**	**0.881**	**0.173**	107/5000
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 50				
Log Risk Aversion Mean: $\mu = 0.2$	**0.137**	**0.909**	**0.184**	57/5000
Cross-sectional Sample Size: 10,000, Time-series Sample Size: 100				
Log Risk Aversion Mean: $\mu = 0.2$	**0.153**	**0.925**	**0.184**	39/5000
Cross-sectional Sample Size: 50,000, Time-series Sample Size: 25				
Log Risk Aversion Mean: $\mu = 0.2$	**0.078**	**0.865**	**0.176**	110/5000
Cross-sectional Sample Size: 50,000, Time-series Sample Size: 50				
Log Risk Aversion Mean: $\mu = 0.2$	**0.141**	**0.893**	**0.187**	56/5000
Cross-sectional Sample Size: 50,000, Time-series5Sample Size: 100				
Log Risk Aversion Mean: $\mu = 0.2$	**0.154**	**0.908**	**0.188**	43/5000

Notes: This table reports the Monte Carlo results for the correct model obtained using our proposed estimation method. They are derived by simulating the general equilibrium model 5000 times. The second column reports the average estimated parameter, where the average is computed over the 5000 simulations, when we use all the Monte Carlo runs and set $\Theta_t = 0.99$ in all cases in which $\Theta_t \geq 1$. Column 3 reports the corresponding coverage probability of a confidence interval with 90% nominal coverage probability. Columns 4 reports the average estimated parameter when we drop all simulations for which $\Theta_t \geq 1$. Column 5 reports the number of case in which $\Theta_t \geq 1$.

Table 7: Monte Carlo Results, Parameter Estimates For Misspecified Model

	Censored Results		*Truncated Results*	
True Parameter	Estimate	Bias	Estimate	N. Cases
Cross-sectional Sample Size: 2,500				
Log Risk Aversion Mean: $\mu = 0.2$	**-0.923**	**-1.123**	**1.224**	1853/5000
Cross-sectional Sample Size: 5,000				
Log Risk Aversion Mean: $\mu = 0.2$	**-0.932**	**-1.132**	**1.224**	1861/5000
Cross-sectional Sample Size: 10,000				
Log Risk Aversion Mean: $\mu = 0.2$	**-0.935**	**-1.135**	**1.227**	1865/5000
Cross-sectional Sample Size: 50,000				
Log Risk Aversion Mean: $\mu = 0.2$	**-0.937**	**-1.137**	**1.229**	1867/5000

Notes: This table reports the Monte Carlo results for the misspecified model obtained using only cross-sectional variation. They are derived by simulating the general equilibrium model 5000 times. The second column reports the average estimated parameter, where the average is computed over the 5000 simulations, when we use all the Monte Carlo runs and set $\Theta_t = 0.99$ in all cases in which $\Theta_t \geq 1$. Column 3 reports the corresponding coverage probability of a confidence interval with 90% nominal coverage probability. Columns 4 reports the average estimated parameter when we drop all simulations for which $\Theta_t \geq 1$. Column 5 reports the number of case in which $\Theta_t \geq 1$.